

Strategien der Geltungssicherung in der qualitativen Sozialforschung: zur Validitätsproblematik im interpretativen Paradigma

Kelle, Udo; Kluge, Susann; Prein, Gerald

Veröffentlichungsversion / Published Version

Arbeitspapier / working paper

Empfohlene Zitierung / Suggested Citation:

Kelle, U., Kluge, S., & Prein, G. (1993). *Strategien der Geltungssicherung in der qualitativen Sozialforschung: zur Validitätsproblematik im interpretativen Paradigma*. (Arbeitspapier / Sfb 186, 24). Bremen: Universität Bremen, SFB 186 Statuspassagen und Risikolagen im Lebensverlauf. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-14266>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC-ND Lizenz (Namensnennung-Nicht-kommerziell-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>

Terms of use:

This document is made available under a CC BY-NC-ND Licence (Attribution-Non Commercial-NoDerivatives). For more information see:

<https://creativecommons.org/licenses/by-nc-nd/4.0>

Strategien der Geltungssicherung
in der qualitativen Sozialforschung.
Zur Validitätsproblematik
im interpretativen Paradigma.

von
Udo Kelle, Susann Kluge und Gerald Prein

Arbeitspapier Nr. 24

Herausgeber: Der Vorstand des Sfb 186
Bremen 1993

Udo Kelle
Susann Kluge
Gerald Prein
Universität Bremen
Sonderforschungsbereich 186
Bereich Methoden und EDV
Wiener Str. - FVG-West
28359 Bremen
Tel.: x49 421 218-4168/-4169
FAX: x49 421 218-4153
Email: ukelle@sfb186.uni-bremen.de
skluge@sfb186.uni-bremen.de
gprein@sfb186.uni-bremen.de

Inhaltsverzeichnis

	Vorwort.....	3
1	Das Problem: Können qualitative Forschungsergebnisse hinreichend valide sein?.....	5
2	Fragen nach Gütekriterien und Standards für die qualitative Sozialforschung	15
2.1	Benötigt die qualitative Forschung überhaupt Gütekriterien?	16
2.2	Benötigt die qualitative Forschung eigene, alternative Gütekriterien? 19	
2.2.1	"Glaubwürdigkeit" statt "interner Validität"?	20
2.2.2	"Übertragbarkeit" bzw. "Anwendbarkeit" statt "externer Validität"	24
2.2.3	Das Risiko der "rhetorischen Validierung"	27
2.3	Müssen die klassischen Gütekriterien den Anforderungen der qualitativen Forschung angepaßt werden?	29
3	Klassische Validitätskonzepte der quantitativen Sozialforschung	35
3.1	Guttman 35	
3.2	Zetterberg 36	
3.3	Cook und Campbell	39
3.4	Der "fallibilistische Kern" klassischer Validitätskonzepte.....	41
4	Methodenprobleme und Validierungsstrategien qualitativer Sozialforschung	43
4.1	Verfahren zur Sicherung der internen Validität	44
4.1.1	Bedrohungen der internen Validität in der qualitativen Sozialforschung 44	
4.1.2	Diskursive Validierung und nominale Gruppentechnik.....	49
4.1.3	Die empirische Überprüfung von Deutungshypothesen am Material	52
4.1.3.1	Sequentielle Analyse.....	52
4.1.3.2	Qualitative Kodierung und Hypothesenprüfung	54
4.2	Verfahren zur Sicherung der externen Validität	59
4.2.1	Bedrohungen der externen Validität in der qualitativen Sozialforschung 59	
4.2.2	Sicherung der Stichprobenvalidität.....	62
4.2.2.1	Das Postulat der Verzerrungsfreiheit und das Prinzip der kriteriengeleiteten Stichprobenziehung.....	62
4.2.2.2	Theoretical sampling, geschichtete Stichproben und "Matrjoschka-sampling"	64
4.2.3	Kommunikative Validierung und qualitative Panels	69
4.2.4	Das Konzept der "Triangulation"	72
5	Literatur	75

Vorwort

Biographie- und Lebenslaufforschung, die den Sichtweisen, Deutungsmustern und Handlungsorientierungen der Akteure eine zentrale Bedeutung für die Theoriebildung zumißt, kann auf qualitative Erhebungs- und Auswertungsmethoden nicht verzichten. Die Zuverlässigkeit und Generalisierbarkeit qualitativer Forschungsergebnisse wird jedoch in der methodologischen Debatte oftmals bezweifelt. Am Sonderforschungsbereich 186 wurde im Sommer 1992 ein Arbeitskreis ins Leben gerufen, der sich schwerpunktmäßig mit Fragen der Validität qualitativer Forschungsmethoden und -ergebnisse befaßt hat. In dem vorliegenden Arbeitspapier werden Diskussionsergebnisse dieses Arbeitskreises dargestellt: hierzu werden verschiedene Beiträge zu diesem Thema aus der Literatur und die in einigen Teilprojekten des Sfb angewendeten Validierungsstrategien diskutiert.

Im ersten Teil des Aufsatzes geht es um angemessene "Standards" oder "Gütekriterien" für qualitative Forschungsmethoden. Hierzu werden verschiedene in der jüngeren Zeit entwickelte Konzepte vorgestellt und auf dieser Basis ein Modell "interner" und "externer Validität" im Kontext qualitativer Forschung entwickelt, dessen Ausgangspunkt spezifische Erkenntnisprobleme und Fehlerquellen qualitativer Forschung bilden.

Im zweiten Teil werden Validierungsstrategien vorgestellt, die der forschungspraktischen Bearbeitung der dargestellten Fehlerquellen und "threats for validity" dienen. Als Möglichkeiten zur Absicherung der "internen Validität" werden diskutiert: die Kombination der Methoden der *analytischen Induktion* und des *permanenten Vergleichs* mit Verfahren der diskursiven Geltungssicherung und EDV-gestützten Verfahren der Textdatenverwaltung. Außerdem werden einige in der qualitativen Sozialforschung oftmals vernachlässigte Aspekte "externer Validität" angesprochen, insbesondere die der Gewinnung und Validierung. Abschließend wird zu der in jüngerer Zeit verstärkt geführten Debatte um die Bedeutung der "Triangulation" von Methoden und Datenquellen aus der Sicht des Sfb Stellung genommen.

Prof. Dr. Walter R. Heinz
Sprecher des Sfb 186

1

2 Das Problem: Können qualitative Forschungsergebnisse hinreichend valide sein?

Seit den Anfängen qualitativer Forschung ist deren Wissenschaftlichkeit und die Gültigkeit und Zuverlässigkeit ihrer Ergebnisse umstritten. Eine Debatte hierüber zwischen Befürwortern und Skeptikern qualitativer Methoden fand jedoch nur äußerst selten statt; in der Regel wurden über die Validität qualitativer Erhebungs- und Auswertungsverfahren zwei getrennte methodologische Diskurse geführt, die kaum aufeinander Bezug nahmen:

Die Argumente der Kritiker und Skeptiker konzentrierten sich sowohl auf die Intersubjektivität, Standardisierbarkeit und Replizierbarkeit des Vorgehens als auch auf Fragen der Stichprobenauswahl - qualitative Forschungsergebnisse wurden einerseits als unzuverlässig betrachtet, weil zu sehr der subjektiven Willkür des einzelnen Forschers unterworfen, und andererseits für nicht verallgemeinerbar, weil die Auswahl der Untersuchungsgruppe nicht durch ein geregeltes Verfahren der Stichprobenziehung zustande gekommen sei. Diese Vorbehalte werden in ähnlicher Form seit den zwanziger Jahren bis in die heutige Zeit vorgetragen - als beispielhaft kann etwa die von Lundberg 1929 geäußerte Kritik an der Untersuchung von Thomas und Znaniecki über den *polish peasant* gelten:

"The scientific value of all these (studies) depends, of course, upon the *validity of the subjective interpretations* of the authors as well as the extent to which the *cases selected are typical*¹. Neither the validity of the sample nor of the interpretations are objectively de-monstrable on account of the informality of the method."
(LUNDBERG 1942, S.169)

Nach Ansicht Lundbergs ist die von Thomas und Znaniecki verwendete Methode der qualitativen Dokumentenanalyse zu informell, um verschiedenen Untersuchern übereinstimmende Urteile und Interpretationen zu erlauben. Welche Schlußfolgerungen aus dem Datenmaterial zu ziehen seien, sei vielmehr in das Belieben des Interpreten gestellt. Zu einem ähnlichen Schluß gelangt auch Bain in seiner Kritik qualitativer Methoden. Qualitative *life histories* könnten nur dann eine valide Basis für theoretische Verallgemeinerungen abgeben, wenn sie Materialien umfaßten, die "*are clearly enough defined and frequent enough in occurrence so that a number of competent observers, working independently, can arrive at like conclusions both as to existence and meaning of the defined data.*" (BAIN 1929, S.155 f.) Da die meisten qualitativen Untersuchungen von solchen methodologischen Standards jedoch weit entfernt seien, sei es evident "*that most so-*

¹ Hervorhebung durch die Verfasser.

called 'scientific' results from the use of life documents, life stories, interviews, diaries, autobiographies, letters, journals etc. are pure poppy-cock." (ebd.)

Das Mißtrauen gegenüber qualitativen Forschern, sie würden auf der Grundlage einer unsicheren Datenbasis mit Hilfe eines methodisch kaum kontrollierten, subjektiven, mehr oder weniger "impressionistischen" Verfahrens Verallgemeinerungen vornehmen, hat den in den vierziger und fünfziger Jahren stattgefundenen Niedergang der qualitativen Forschungstradition wesentlich mitverursacht (vgl. BERTAUX 1981, S.5). Bis heute findet es seinen Niederschlag darin, daß qualitative Verfahren in den meisten Standardwerken über Methoden empirischer Sozialforschung weitgehend ausgeklammert bleiben², weil i.a. davon ausgegangen wird, daß ihre *"Zuverlässigkeit und Gültigkeit nicht abzuschätzen"* sei (MAYNTZ, HOLM, HÜBNER 1969, S.92) oder daß für sie keinerlei methodische Standards und Gütekriterien *"jenseits subjektiver Evidenzerlebnisse formulier- und überprüfbar"* seien (SCHNELL, HILL, ESSER 1989, S.110), wobei solche Urteile in der Regel nicht mehr begründet werden, sondern sich teilweise zu einer Art orthodoxen Lehrmeinung verfestigt haben. Die Kritik an qualitativen Methoden beschränkt sich dabei zumeist, wie bei Schnell und Kollegen, auf wenige Sätze, in denen die mangelnde Validität und Reliabilität qualitativen Vorgehens nur noch apodiktisch festgestellt, aber nicht mehr systematisch begründet wird³. Qualitative Forschungsmethoden haben sich deshalb seit ihrer Renaissance in den siebziger Jahren weitgehend abseits vom "Mainstream" sozialwissenschaftlicher Methodologie entwickeln müssen. Wie tief das Schisma zwischen qualitativen und quantitativen methodologischen Traditionen nach wie vor geht, wird daran deutlich, daß es für den Bereich qualitativer Forschung mittlerweile eigene Lehr- und Handbücher gibt (LAMNEK 1988; SPÖHRING 1989; FLICK u.a. 1991, MAYRING 1990) und daß kaum ein sozialwissenschaftlicher Methodiker gleichzeitig über qualitative und quantitative Verfahren publiziert.

² Das Handbuch "Methoden der empirischen Sozialforschung" (Schnell, Hill, Esser 1989) enthält gerade drei Seiten zum offenen Leitfadengespräch und zwei Seiten zum narrativen Interview, während Methoden des Standardisierten Interviews rund 60 Seiten einnehmen. Verfahren der nicht-standardisierten Teilnehmenden Beobachtung werden in diesem Standardwerk mit wenigen Sätzen abgehandelt. Im Bereich der Inhaltsanalyse wird die Existenz einer "qualitativen Inhaltsanalyse" (Mayring 1985, 1988) ebenso wie in Kromreys Monographie "Empirische Sozialforschung" (1990) gar nicht erst erwähnt. Hinsichtlich der Beobachtungsmethoden in der Sozialforschung erwähnt Kromrey zwar, daß es Methoden der offenen teilnehmenden Beobachtung gibt (1990, S.189), im weiteren werden dann jedoch nur "systematische Beobachtungsmethoden" dargestellt. Diese erfordern nach Kromrey aus stichprobentheoretischen Erwägungen die genaue definitorische Festlegung der zu beobachtenden Situationen, der Klasse der relevanten Ereignisse, der Beobachtungszeiträume und -orte und die Aufstellung eines Kategorienschemas. Die Darstellung nicht-standardisierter Interviewformen nimmt einen Absatz ein (1990, S.211).

³ Nur Opp kritisiert an mehreren Stellen ausführlich (1976, 1987) den Anspruch der qualitativen Sozialforschung, theoriefindende Methoden zu formulieren.

Daß die Auffassung von der Unzuverlässigkeit qualitativer Forschungsmethoden sich innerhalb des "Mainstreams" in der geschilderten Weise verfestigt hat, ist sicherlich zu einem großen Teil darauf zurückzuführen, daß eine Antwort auf die dargestellten Vorbehalte und Kritikpunkte von Seiten qualitativer Forscher lange Zeit ausgeblieben ist. Zwar spielte bei Wiederbelebung der qualitativen Tradition in der Sozialforschung das Validitätsargument eine entscheidende Rolle, nur wurde hier Validität in ganz anderer Weise definiert: als größere Gegenstandsadäquatheit qualitativer Methoden und höhere Relevanz qualitativer Forschungsergebnisse. Dies wurde weitgehend nicht methodologisch, sondern handlungstheoretisch begründet: für die *interpretative Soziologie* (GIDDENS 1984) bzw. das *interpretative Paradigma* (WILSON 1981) ist soziales Handeln grundsätzlich nicht erklärbar ohne einen Rekurs auf die Deutungsmuster, Sichtweisen und Alltagswissensbestände der handelnden Akteure selber. Ähnlich wie ein Ethnologe hat es auch der soziologische Untersucher dabei beständig mit sozialen Lebens- und Praxisformen zu tun, deren Wissens- und Regelbestände ihm nicht von vornherein verfügbar sind. Wendet er nun eine strikt quantitative Methodologie an, so greift er bei der Theoriebildung quasi ins Leere, denn die Festlegung der relevanten Untersuchungsdimensionen und die Definition und Operationalisierung von Variablen muß dann *vor jeglichem Kontakt mit dem empirischen Feld* erfolgen. Jenes Wissen, welches der Untersucher für die Formulierung seiner Hypothesen benötigt, kann er sich aber nur durch den Kontakt mit den untersuchten sozialen Lebenswelten, d.h. seinem empirischen Feld verschaffen. Die implizite Vorstellung, von der die quantitative und hypothetiko-deduktive Methodologie ausgehe, daß nämlich *"die Forscher (...) in genügendem Maße inhaltsreiche Vorstellungen über die untersuchten Wirklichkeitsbereiche zur Verfügung hätten"* (Gerdes 1979, S.5), erweise sich deshalb als zentrales Hindernis für den sozialwissenschaftlichen Forschungsprozeß: *"Sachverhalte, über die der Forscher keine Vorstellungen hat, weil er den betreffenden Wirklichkeitsbereich nicht umfassend kennt, können nämlich in seinen Hypothesen gar nicht auftauchen, werden also auch nicht getestet und fehlen folglich im wissenschaftlichen Bild dieses Wirklichkeitsbereichs. Sind solche Sachverhalte konstitutiv für den untersuchten Bereich, bleibt die wissenschaftliche Darstellung ohne ausreichenden Bezug zur Wirklichkeit - und zwar selbst dann, wenn sie sich ausschließlich auf empirisch bestätigte Hypothesen stützen kann."* (ebd., S.5)

Der in der Soziologie der sechziger und siebziger Jahre vorherrschende Trend zu hypothetiko-deduktiven und quantitativen Forschungsdesigns, der durch die Standardisierung der Datenerhebung und -auswertung eine hohe Zuverlässigkeit und Objektivität der Forschungsergebnisse garantieren und durch die Ziehung umfangreicher Zufallsstichproben deren Generalisierbarkeit sicherstellen sollte, wurde von Vertretern des interpretativen Paradigmas dafür verantwortlich gemacht, daß sich in der Soziologie *"ein verminderte(s)"*

Verständnis der empirischen sozialen Welt" und eine *"künstliche Auffassung von Realität"* (FILSTEAD 1979, S.31) durchgesetzt hätte. Der quantitativ forschende Soziologe käme in seinem Forschungsalltag nur noch *"selten in Berührung mit dem, was er zu verstehen"* versuche (ebd., S.29), und er vernachlässige über der Beschäftigung mit methodisch-statistischen Detailfragen die theoretische Relevanz und empirische Bedeutsamkeit seiner Forschungsfragen und -ergebnisse.

Die Kontroverse um die Validität qualitativer Forschungsmethoden war lange Zeit auf diese Weise in zwei getrennte Diskurse aufgespalten: Die Kritiker qualitativer Forschungsmethoden monierten die *mangelhafte Validität* der hiermit gefundenen Ergebnisse, weil sie einer unkontrollierten Erhebungs- und Auswertungsprozedur entstammen und deswegen nicht replizierbar, nicht intersubjektiv überprüfbar und damit nicht *reliabel*, sondern in hohem Grade von der Person des Forschers abhängig seien. Zudem seien qualitative Forschungsergebnisse nicht *generalisierbar*, weil kein geregeltes Sampling-Verfahren zur Anwendung komme und deswegen nicht gesichert sei, daß die untersuchten Fälle in irgendeiner Weise "typisch" oder "charakteristisch" für den untersuchten Gegenstandsbereich seien. Von den Vertretern des interpretativen Paradigmas wurde demgegenüber die *hohe Validität* qualitativer Forschungsmethoden betont, weil diese einen direkteren Zugang zu der erforschten sozialen Realität ermöglichen und damit *relevantere* Forschungsergebnisse erbringen könnten als umfangreiche Erhebungen mit standardisierten Instrumenten. Dem Begriff Validität wurde dabei von beiden "Lagern" eine unterschiedliche Bedeutung gegeben: auf Seiten des hypothetiko-deduktiven Paradigmas wurde hierunter im wesentlichen *Reliabilität*, d.h. Replizierbarkeit, Objektivität und Stabilität der Forschungsergebnisse, und *Generalisierbarkeit*, d.h. vor allem Repräsentativität, verstanden. Qualitative Methodologen verstanden hierunter vor allem *Relevanz* und *Nähe zum Gegenstand*, die sich darin ausdrücken sollten, daß der Untersucher versucht, die im Forschungsfeld bedeutsamen sozialen Regeln und sozialen Wissensbestände exploratorisch zu erkunden. Zwei vollständig verschiedene "Validitätssprachspiele" wurden etabliert, so daß die Kritik der jeweils "ande-ren Seite" kaum zur Kenntnis genommen wurde:

- * Vorwiegend quantitativ orientierte Soziologen räumen zwar oftmals ein, daß für die Formulierung sinnvoller Hypothesen ein vorheriger Feldzugang erforderlich ist (bspw. BARTON, LAZARSFELD 1984; FRIEDRICHS 1980). Solche Überlegungen wurden bereits in den fünfziger Jahren angestellt⁴, ohne daß hieraus die notwendige Konsequenz gezogen

⁴ Der Aufsatz von Barton und Lazarsfeld wurde erstmals 1955 veröffentlicht.

wurde, eine *Methodologie exploratorischer Forschung* zu entwickeln. Stattdessen findet sich oft die Auffassung, daß für eine empirisch begründete Hypothesengenerierung ein mehr oder weniger impressionistisches Verfahren ausreichend sei (FRIEDRICHS 1980, S. 52), wobei die seit den fünfziger Jahren stattgefundene Entwicklung methodisch kontrollierter qualitativer Verfahren ignoriert⁵ wird. Hinzu kommt ein bedeutsames logisches Problem: wenn, wie etwa Mayntz, Holm und Hübner durchaus zu Recht schreiben, die Verwendung von unsystematischen Verfahren, bei denen der "*Prozeß der Datenermittlung und Dateninterpretation nicht standardisiert ist*" (1969, S.93) zu einer Beliebigkeit von Interpretationen führen kann, bleibt unverständlich, wieso dieselben Autoren überhaupt solche Verfahren zur Hypothesengenerierung empfehlen. Die Entwicklung *beliebiger Hypothesen* kann kaum als sinnvolles Ziel von explorativen Vorstudien betrachtet werden. Wenn allerdings zumindest implizit davon ausgegangen wird, daß qualitativen Verfahren wenigstens ein Mindestmaß an Rationalität und Systematik zu eigen ist, die zur Formulierung rational begründeter, brauchbarer Hypothesen führt, so müßten diese Systematik und Rationalität expliziert und weiterentwickelt werden. Vorschläge zur Systematisierung qualitativer Verfahren von Seiten quantitativ orientierter Soziologen beziehen sich allerdings meist nur darauf, daß exploratorische Verfahren standardisiert werden müßten, um zuverlässige und gültige Interpretationen zu erbringen (MAYNTZ, HOLM, HÜBNER 1969, S.93; FRIEDRICHS, LÜDTKE 1973, S.33). Solche Vorschläge sind zur Entwicklung einer brauchbaren Methodologie empirisch begründeter Hypothesengenerierung kaum brauchbar, denn die Standardisierung von exploratorischen Verfahren (etwa durch die Festlegung eines Beobachtungsschemas und die Operationalisierung von Beobachtungskategorien, wie von Friedrichs und Lüdtke gefordert) verlangt, daß die relevanten Untersuchungsgegenstände *vor der Untersuchung* spezifiziert werden und Hypothesen über Zusammenhänge zwischen diesen Untersuchungsgegenständen formuliert werden. Die Standardisierung exploratorischer Verfahren führt dazu, daß Hypothesen nicht mehr generiert werden können, sondern vorher formulierte Hypothesen überprüft werden und der exploratorische Charakter des Vorgehens damit verloren geht. Interessanterweise wird i.d.R. selbst dann, wenn die handlungstheoretische Kritik des interpretativen Paradigmas an der konventionellen Variablensoziologie geteilt wird, die Notwendigkeit eines exploratorischen und qualita-

⁵ Bspw. solche Verfahren wie die *analytische Induktion* (Cressey 1971, Lindesmith 1968), die *grounded theory* (Glaser, Strauss, Corbin), die *objektive Hermeneutik* (Oevermann u.a. 1979) oder die von Gerhardt vorgestellte, auf den Arbeiten von Weber fußende Methode der *typisierenden Prozeßstrukturanalyse* (Gerhardt 1985, 1991).

tiven Vorgehens kaum thematisiert und methodologische Regeln hierfür nicht expliziert: So betont Esser, der die handlungstheoretischen Defizite der Variablensoziologie deutlich herausstellt (ESSER 1989), welche die individuellen Deutungsmuster, Handlungsorientierungen und Entscheidungsleistungen bei der Formulierung soziologischer Handlungserklärungen nicht genügend berücksichtigt, zwar die Notwendigkeit, abstrakte Modellierungen auf der Basis *induktiv gewonnener Beobachtungen* (ESSER 1991, S.76) vorzunehmen, vernachlässigt jedoch qualitative Forschungsmethoden fast vollständig. Während er verstehende Methoden in der Sozialforschung gleichsetzt mit "Operationalisierung" und "Datenerhebung" (ESSER 1991, S.92), widmet er der Frage, auf welche (methodisch kontrollierte) Weise denn Beobachtungen induktiv zu abstrakten Modellierungen führen könnten allenfalls den Halbsatz, hierzu sei intensive "Teilnahme" am Forschungsgegenstand notwendig, die sowohl bei qualitativen als auch bei quantitativen Methoden durchgeführt werden könnte (ESSER 1991, S.92). Auch dort, wo die handlungstheoretische Begründung für das "Relevanz-argument" des interpretativen Paradigmas akzeptiert wird, werden hieraus methodologische Konsequenzen nicht gezogen.

- * Auch in der qualitativen Forschung wurde die aus dem methodologischen "Mainstream" geäußerte Skepsis gegenüber der Validität qualitativer Forschung, die ihren Ausdruck im Reliabilitäts- und Repräsentativitätsargument fand, vielfach ignoriert. Demgegenüber war lange Zeit die Tendenz feststellbar, Ergebnisse qualitativer Forschung allein deswegen für valide zu halten, weil sie durch einen *direkten Kontakt mit der untersuchten Realität* zustande gekommen seien (wobei der Forderung nach Reliabilität des Vorgehens allenfalls Lippentribut gezollt wird, ohne daß hierzu konkrete Vorschläge gemacht werden (vgl. etwa FILSTEAD 1979, S.36)). Ihren methodologischen und wissenschaftstheoretischen Ausdruck kann diese Position in einem offensiv vertretenen Induktivismus finden, wie er etwa von Glaser und Strauss in ihrem 1967 erschienenen Standardwerk zur qualitativen Methode geäußert wird:

"To generate theory (...) we suggest as the best approach an initial, systematic discovery of the theory from the data of social research. Then one can be relatively sure that the theory will fit and work." (GLASER, STRAUSS 1967, S.3)

Nach dieser Vorstellung soll der Forscher die theoretische Literatur über den Gegenstandsbereich vollständig ignorieren, um sicherzustellen, daß das *Emergieren* von

Kategorien aus dem Datenmaterial nicht behindert wird (ebd., S.37). Interessanterweise sind solche induktivistischen *tabula rasa* Modelle menschlicher Forschungstätigkeit⁶, wonach dem Untersucher theoretische Konzepte und Verallgemeinerungen quasi von selber zuwachsen, wenn er sich seinem Gegenstandsbereich nur vorurteilsfrei und theorielos nähert, als eine der wesentlichen Wurzeln des *logischen Positivismus* zu betrachten (POPPER 1989, S.9ff.), eine erkenntnistheoretische Richtung, die in qualitativen Forscherkreisen einen ansonsten eher schlechten Ruf genießt. Daß induktivistische Konzeptionen des Forschungsprozesses auf einer naiven Fehleinschätzung menschlichen Erkenntnisvermögens beruhen und damit eine unhaltbare Vereinfachung des Forschungsprozesses darstellen, hat die moderne Wissenschaftstheorie zur Genüge aufgezeigt: die Realität kann stets nur durch die "Linsen" eines konzeptuellen Netzwerkes betrachtet werden:

"Both historical examples and recent philosophical analysis have made it clear that the world is always perceived through the 'lenses' of some conceptual network or other and that such networks and the languages in which they are embedded may, for all we know, provide an ineliminable 'tint' to what we perceive." (LAUDAN 1977, S.15)

Nicht-induktivistische Modelle des qualitativen Forschungsprozesses müssen demgegenüber davon ausgehen, daß theoretische Konzepte nicht aus dem Datenmaterial "emergieren", sondern daß der Untersucher die im Feld gesammelten Daten *auf der Grundlage seines theoretischen Vorwissens* deutet und interpretiert und sich um eine Erklärung der so gesammelten Daten bemüht. Hierbei läßt sich die Frage nach der Validität qualitativen Vorgehens nicht mehr induktivistisch (d.h. mit dem Hinweis auf den direkten Feldzugang) beantworten. Seit dem Beginn der achtziger Jahre wurden eine Reihe von Diskussionsbeiträgen zum Gültigkeitsproblem in der qualitativen Sozialforschung vorgelegt, die teilweise auf einer nicht-induktivistischen Konzeption des qualitativen Forschungsprozesses beruhen (TERHART 1981; HEINZE, THIEMANN 1982; KLEINING 1982; VOLMERG 1983; KÜCHLER 1983; GERHARDT 1985; LINCOLN, GUBA 1985, S.290 ff.; KIRK, MILLER 1986; FLICK 1987, 1993; LAMNEK 1988; KVALE 1989, 1991; LECOMPTE, PREISSLE 1993, S. 315 ff. etc.). Diese Beiträge beruhen auf jeweils

⁶ Auch Glaser und Strauss räumen ein, daß der Forscher sich nicht als *tabula rasa* der Realität nähern kann, geben aber in ihrer 1967 erschienenen Monographie keinen Hinweis darauf, wie der Forscher sein tatsächlich vorhandenes Vorwissen im Untersuchungs-prozeß sinnvoll einsetzen kann.

verschiedenen methodologischen und erkenntnistheoretischen Grundlagen und beziehen sich auf jeweils verschiedene Validitätsprobleme. Manche Autoren bemühen sich dabei um eine generelle Klärung der Validitätsproblematik im interpretativen Paradigma, andere beleuchten nur einzelne Aspekte oder schlagen spezifische Validierungsstrategien vor. Allerdings zeichnet sich in der Diskussion bislang kein ähnlicher Konsens über die Bedeutung von Konzepten wie Validität, Reliabilität oder Objektivität für qualitative Forschungsmethoden ab, wie er in der quantitativen Sozialforschung - insbesondere aufgrund der Arbeiten von GUTTMAN (1941) oder von COOK und CAMPBELL (1979) mittlerweile erreicht worden ist.

Bevor konkrete Strategien und Methoden der Validierung qualitativer Forschungsergebnisse im einzelnen dargestellt und auf der Basis forschungspraktischer Erfahrungen von Teilprojekten des Sfb evaluiert werden können, soll deshalb im folgenden eine *Klärung des Validitätsbegriffs im Kontext qualitativer Sozialforschung* versucht werden.

2 Fragen nach Gütekriterien und Standards für die qualitative Sozialforschung

In der bereits erwähnten Literatur über Validität in der qualitativen Sozialforschung wird regelmäßig die Auffassung geäußert, daß sich die "traditionellen" oder "konventionellen" Gütekriterien und Standards aus dem Kontext quantitativer Sozialforschung nicht ohne weiteres auf qualitative Erhebungs- und Auswertungsstrategien übertragen ließen (vgl. AGAR 1986; CORBIN, STRAUSS 1990, S.418; KIRK, MILLER 1986, S.14; SAHNER 1989; FLICK 1987, S.247; LINCOLN, GUBA 1985, S.293 ff.). Darüber, daß es offensichtlich sinnlos ist, die in der Testpsychologie und experimentellen Forschung entwickelten Kriterien und Konzepte ("split-half-reliability") in derselben Form in qualitativen Forschungsdesigns einzubeziehen, herrscht unter den Diskutanten weitgehender Konsens. Allerdings läßt diese Feststellung drei völlig verschiedene Standpunkte zu:

1. Jede Form der Bewertung von qualitativer Forschungspraxis anhand allgemeiner Gütekriterien oder Standards wird zurückgewiesen.
2. Es werden eigene Gütekriterien und Standards für die qualitative Sozialforschung formuliert.
3. Die für quantitative Methoden entwickelten Gütekriterien und Standards werden für die Bedürfnisse der qualitativen Forschung "redefiniert" bzw. angepaßt.

Wir wollen diese drei Standpunkte, die in der Literatur von jeweils verschiedenen Autoren vertreten werden, im folgenden im Detail diskutieren.

2.1 Benötigt die qualitative Forschung überhaupt Gütekriterien?

"We don't do reliability and validity", schreibt Smith (1984) und gibt zum Ausdruck, daß die philosophischen Grundlagen des interpretativen Paradigmas generell die Anwendung von Gütekriterien und Standards obsolet erscheinen lassen. Grundlage solcher Argumentationen ist in der Regel die Annahme, daß qualitative Forschung einem eigenen "Paradigma" mit eigenen "Wahrheitsansprüchen" entstamme (z.B. LINCOLN, GUBA 1985, S.35 ff.), oftmals wird hierbei auch auf die von Dilthey (1924) getroffene Unterscheidung zwischen "Erklären" und "Verste-hen" zurückgegriffen (LAMNEK 1988, S.67 ff.). Gemäß dieser Unterscheidung würde qualitative Sozialforschung auf einer eigenen *Kunstlehre des Verstehens* beruhen, die von der experimentellen Methode der "erklärenden" Naturwissenschaften streng zu unterscheiden sei. An dieser Stelle können nicht die Grundlagen der "Erklären:Verstehen-Kontroverse" diskutiert werden, eine ausführliche Dar-stellung findet sich bspw. bei Apel (1979) oder bei Kelle (1992). Es soll vielmehr aufgezeigt werden, daß sich die Frage nach Gütekriterien auch dann stellt, wenn man die Annahme akzeptiert, daß ein grundlegender Dualismus zwischen experi-menteller Methode und hermeneutischer Kunstlehre des Verstehens besteht.

Hierzu ist eine Klärung des Begriffs "Kunstlehre" erforderlich. Hiermit kann etwa die Aufforderung an den Forscher gemeint sein, sich wie ein Künstler zu verhalten, wie dies etwa Eisner (1981) tut: der "künstlerisch orientierte Forscher" will, anstatt die Wahrheit zu suchen, Bilder erschaffen, "*images that people will find meaningful and from which their fallible and tentative views of the world can be altered, rejected or made more secure*" (EISNER 1981, S.9). Für Kunstwerke wird kein Wahrheitsanspruch erhoben; auch Beschreibungen von Ereignissen, die nie stattgefunden haben oder Abbildungen von nicht existierenden Gegenständen können einen hohen künstlerischen Wert besitzen. Nur würde die Übertragung solcher Kriterien auf empirische Forschung offensichtlich absurde Ergebnisse haben, weil demnach "Gulliver's Reisen" dieselbe wissenschaftliche Qualität besitzen müßten wie ein ethnographischer Forschungsbericht. Eine solch weitgehende Konsequenz zieht auch Eisner offensichtlich nicht, sonst würde er kaum davon sprechen, daß die "Bilder" des Forscher-Künstlers die "fehlbaren" ("falliblen") Ansichten der Menschen korrigieren könnten: fehlbare Ansichten können nur durch die Darstellung von Sachverhalten korrigiert werden und der Forscher kann anders als der Künstler nicht reine Fiktionen präsentieren, sondern muß sich um eine

Beschreibung sozialer Realität bemühen. Begriffe von "falsch" und "richtig" spielen im Kontext künstlerischer Betätigung keine Rolle, wohl aber für die empirische Sozialforschung: während es für die Beurteilung eines Kunstwerks zumeist ohne Belang ist, ob die dargestellten Ereignisse, Gegenstände etc. tatsächlich existieren (ob etwa der von Rilke beschriebene Panther im "Jardin des Plantes" je gelebt hat oder ob im Moskau des vorigen Jahrhunderts tatsächlich ein Mann namens Rodian Raskolnikoff seine Vermieterin umgebracht hat), ist die Frage der Realitätsadäquatheit für die Sozialforschung von zentraler Bedeutung. Tatsächlich taucht das Problem einer fehlerhaften Beschreibung sozialer Phänomene relativ häufig in der ethnographischen Feldforschung und qualitativen Sozialforschung auf:

Kirk und Miller (1986) etwa schildern, wie bei der Untersuchung von Praktiken des Drogenkonsums in Peru zu Beginn der Feldphase aufgrund einer falschen Fragetechnik invalide Daten gesammelt wurden. Die Frage *"Erzählen Sie uns über Koka?"* veranlaßte die Interviewten dazu, ausweichende Antworten zu geben. Erst als sie gefragt wurden *"Wann geben Sie Ihren Tieren Koka?"* oder *"Wie haben Sie herausgefunden, daß sie Koka nicht mögen?"* brachte die Informanten dazu, tatsächlich ihre eigenen Erfahrungen mit der Droge zu berichten.

In einer Studie über Sozialhilfekarrieren am Sonderforschungsbereich 186 wichen Berichte, die ein Teil der Betroffenen in qualitativen Interviews gaben, in vielerlei Hinsicht (auch bezogen auf objektive Daten) von den in den Akten der Sozialämtern vorliegenden Darstellungen ab.

Bekannt ist auch die große Divergenz zwischen ethnologischen Forschungsergebnissen von MEAD (1978) und FREEMAN (1983) über "Jugend und Sexualität" auf Samoa.

In allen drei Fällen können die jeweils divergenten Ergebnisse auf der Basis weiterer theoretischer und methodologischer Annahmen erklärt werden, wobei die Unzuverlässigkeit von Datenquellen deutlich wurde. Stützt sich der Forscher bei seinem Bericht über das Verhalten der jeweiligen Akteure, über die in einer Kultur geltenden Regeln und Normen, allein auf diese Quellen, sind die Ergebnisse zwangsläufig fehlerhaft:

Die anfänglichen Auskünfte der von Kirk und Miller befragten Peruaner über ihren Koka-Konsum waren offensichtlich falsch. Diese fehlerhaften Ergebnisse können etwa mit einer *Theorie sozialer Erwünschtheit* erklärt werden. Im Fall der von Mead und Freeman untersuchten Kultur können divergierende Ergebnisse dadurch erklärt werden, daß beide Untersucher unterschiedliche Informanten befragt hatten und eine unterschiedliche Rolle im Feld einnahmen: Mead erhielt Informationen vor allem von jungen Frauen ihrer eigenen Altersgruppe, zu denen sie besonders leichten Zugang hatte. Hierdurch seien eine Reihe von Verzerrungen ihrer Studie zu erklären, wie sie später einräumte.

Im Gegensatz zu diesen Ergebnissen qualitativer und ethnographischer Forschung können zwei verschiedene künstlerische Darstellungen desselben Gegenstandes oder desselben Sachverhaltes sich grundlegend unterscheiden, ohne daß überhaupt Fragen nach der "Angemessenheit" oder "Richtigkeit" der Darstellung laut werden. Anders gesagt: würden die künstlerischen Kriterien bei der Beurteilung etwa von Gemälden dieselben sein wie sie i.a. an Ergebnisse qualitativer Studien oder ethnographischer Feldforschungen angelegt werden, könnte Picasso nur schwerlich als Künstler bezeichnet werden.

Wird qualitative Forschung als "Kunstlehre" bezeichnet, so kann dies also nicht bedeuten, daß ihre Ergebnisse nur mit ästhetischen oder künstlerischen Maßstäben gemessen werden dürfen. Vielmehr kann sich hierin die methodologische Annahme ausdrücken "*daß ethnographische Beschreibungen methodisch nicht kanonisierbar, in Regeln beschreibbar sind*" (MATT 1992, S.12). Die Tätigkeit des Untersuchers ist demnach in einer solchen Weise individuell und ideosynkratisch, daß ihre Regeln nicht explizierbar sind. Die Frage, bis zu welchem Grad qualitative Forschung regelgeleitet, methodisch kontrolliert verlaufen kann und inwieweit sie ein nicht zu eliminierendes subjektives Moment enthält, können wir an dieser Stelle offenlassen - die Frage danach, inwieweit qualitative Forschungsergebnisse oder ethnographische Beschreibungen die tatsächlichen Regel- und Wissensbestände einer Kultur, die Deutungsmuster und Handlungsorientierungen der Akteure zutreffend und angemessen wiedergeben, ist davon, wie gezeigt wurde, gar nicht berührt.

2.2 Benötigt die qualitative Forschung *eigene, alternative Gütekriterien*?

Während die Auffassung, daß qualitative Forschung als "Kunstlehre" des Verstehens keiner Gütekriterien bedürfe, eine Minderheitenposition (bspw. SMITH 1984, 1990) darstellt, neigt die Mehrheit der bereits erwähnten Autoren der Auffassung zu, daß zwar Gütekriterien in irgendeiner Form notwendig seien, diese jedoch nicht ohne weiteres aus dem Kontext quantitativer und experimenteller Forschung übernommen werden könnten (etwa CORBIN, STRAUSS 1990, S.418; FLICK 1987, S.247; LAMNEK 1988, S.140; LINCOLN, GUBA 1985, S.293). Zumeist bleibt die Argumentation jedoch an diesem Punkt stehen: die "klassischen Gütekriterien" Validität, Reliabilität, Objektivität und Repräsentativität werden diskutiert und anschließend auf deren beschränkte Anwendbarkeit im Rahmen qualitativer Forschung verwiesen (LAMNEK 1988, S.140 ff.). Oft verbleibt die Feststellung, man benötige alternative Gütekriterien auch vollständig im Programmatischen (etwa bei FLICK 1987, S.247). In der Regel werden *Strategien zur Validierung* und nicht Gütekriterien diskutiert - auch wenn, wie bei Lamnek, Verfahren zur Validierung ("kommunikative Validierung") kurzerhand zu qualitativen *Validierungskriterien* erklärt werden. Eigene Gütekriterien für qualitative Sozialforschung werden nur äußerst selten expliziert, so etwa bei Lincoln und Guba (1985) oder bei Corbin und Strauss (1990).

Lincoln und Guba gehen davon aus, daß die konventionellen Gütekriterien nur dann sinnvoll verwendet werden können, wenn die erkenntnistheoretischen Axiome des "*naiven Realismus*" und der "*linearen Kausalität*" akzeptiert werden, die ihrer Auffassung nach quantitativer Methodik zugrundeliegen. Qualitative Forschung gehe dahingegen von grundsätzlich anderen erkenntnistheoretischen Voraussetzungen aus, so daß Standards und Kriterien zur Beurteilung qualitativer Studien entwickelt werden müßten, welche auch terminologisch vollständig unabhängig von dem konventionellen Validitätskonzepten verwendet werden könnten. Dies versuchen sie anhand verschiedener klassischer Gütekriterien zu demonstrieren. Wir wollen im Folgenden die Frage, inwieweit alternative Gütekriterien für die qualitative Sozialforschung notwendig sind, exemplarisch an den von Lincoln und Guba vorgenommenen begrifflichen Gegenüberstellungen zwischen "Glaubwürdigkeit" und "interner Validität" einerseits und "Übertragbarkeit" und "externer Validität" andererseits diskutieren.

2.2.1 "Glaubwürdigkeit" statt "interner Validität"?

Dem Begriff der internen Validität liege, so Lincoln und Guba, ein naiver Realismus zugrunde - die Vorstellung einer Wirklichkeit, welche unabhängig vom Beobachter existiere und die Überprüfung von Hypothesen mit dem Ziel ihrer "Verifikation" oder "Falsifikation" zulasse. Dieser naive Realismus müsse ersetzt werden durch die Annahme vielfältig konstruierter Realitäten:

"When naive realism is replaced by the assumption of multiple constructed realities, there is no ultimate benchmark to which one can turn for justification - whether in principle or by a technical adjustment via the falsification principle. 'Reality' is now a multiple set of mental constructions." (LINCOLN, GUBA 1985, S.295)

Der Verweis auf konstruktivistische, hermeneutische und postmoderne Strömungen der Erkenntnistheorie findet sich häufig auch bei anderen Autoren, die die Bedeutung von Validitätskonzepten für die qualitative Sozialforschung diskutieren. So bemüht sich etwa Kvale (1991) darum, unter expliziter Bezugnahme auf moderne philosophische Ansätze, wo *"die Auffassung der Wahrheit als Korrespondenz zwischen Sätzen und einer unabhängigen objektiven Wirklichkeit in Frage gestellt wurde"* (S.428), die Frage nach der Validität qualitativer Forschungsergebnisse zu beantworten:

"Wissen wird dort nicht als Spiegelbild einer objektiven Natur verstanden und Bewußtsein nicht als Kopie oder Repräsentation einer unabhängigen Welt. Die Unterscheidung zwischen der sozialen Wirklichkeit und ihrer sprachlichen Beschreibung stellt sich als künstlich heraus. Die Sprache erweist sich als konstitutives Merkmal der sozialen Wirklichkeit, deren Charakter sie bestimmt." (KVALE 1991, S.428)

Kvale vernachlässigt allerdings an dieser Stelle die wichtige Unterscheidung zwischen den Konstruktionen der Akteure einer sozialen Lebensform und denen des Forschers, welche etwa Alfred Schütz im Auge hatte, als er zwischen den Theorien 1.Ordnung (den Theorien die die in einer Sozialwelt Lebenden konstruieren) und den Theorien 2. Ordnung (den Theorien, die Forscher aufgrund ihrer Beobachtungen in dieser Sozialwelt formulieren) unterschied. Nur hinsichtlich der Konstruktionen der Laien kann sinnvoll davon gesprochen werden, daß sie die soziale Wirklichkeit mitkonstituieren, während die Konstruktionen der Forscher nur äußerst selten und sehr vermittelt auf die untersuchte Wirklichkeit Einfluß nehmen können: so kann etwa davon gesprochen werden,

daß eine bestimmte von allen Gesellschaftsmitgliedern geteilte Vorstellung von den Pflichten eines Amtsinhabers dieses Amt als soziale Institution erst konstituiert, nicht jedoch davon, daß es erst durch seine Beschreibung in einem Forschungsbericht konstituiert werden würde. Auch wenn man von dem interpretativen Charakter sozialen Handelns ausgeht, d.h. die Auffassung vertritt, daß soziale Strukturen wesentlich durch die Vorstellungen mitkonstituiert werden, welche die Akteure von ihnen haben, kann man nicht auf die (zumindest implizite) Annahme verzichten, daß so etwas wie eine Entsprechung ("Korrespondenz") zwischen sozialen Phänomenen (bspw. dem Alltagswissen, den Deutungsmustern, den Konstruktionen der Akteure) einerseits und den Beschreibungen dieser Phänomene möglich sein muß. Ein radikaler Konstruktivismus dahingegen, bei dem die Vorstellung einer vom Beobachter unabhängigen Realität aufgegeben wird, ist zur erkenntnistheoretischen Begründung von Forschungspraxis ungeeignet, wie Hammersley (1992) ausführt:

"(...) it is a view that can carry radical implications if we apply it to researchers themselves, as we surely must if we are to be consistent. Once we do this, we are led to ask whether research reports are not simply constructions that reflect the nature of the researcher and the research process, rather than representations of reality."
(HAMMERSLEY 1992, S.196)

Weder Lincoln und Guba noch Kvale ziehen allerdings die folgerichtige Konsequenz aus ihren konstruktivistischen Argumentationen, daß wissenschaftliche Beschreibungen sozialer Realität vollständig beliebig seien. Der Konstruktivismus dient bei ihnen (in Verbindung mit einer anti-positivistischen Rhetorik) nur zur erkenntnistheoretischen Legitimation, nicht jedoch zur tatsächlichen Begründung qualitativen Vorgehens. Dabei wird eine "korrespondenztheoretische" bzw. "realistische" Erkenntnistheorie, bei der zwischen erforschter Realität und den Konstruktionen des Forschers unterschieden wird, quasi durch die Hintertür wieder eingeführt. Kvale schlägt unter Rückgriff auf Konzepte der Aktionsforschung die "kommunikative Validierung" (KLÜVER 1979; LECHLER 1982; HEINZE, THIEMANN 1982) der Forschungsergebnisse durch die beforschten Akteure vor. Dieses methodische Verfahren *"sich der Gültigkeit einer Interpretation dadurch zu vergewissern, daß eine Einigung resp. Übereinstimmung über die Interpretation zwischen Interviewten und Interpreten hergestellt wird"*, benötigt allerdings keine spezielle *Konsenstheorie der Wahrheit*, wie Kvale behauptet. Vielmehr kann dieses Vorgehen überhaupt nicht sinnvoll ohne korrespondenztheoretische Annahmen angewendet

werden, d.h. der Forscher muß davon ausgehen, daß eine prinzipielle *Übereinstimmung zwischen der Realität (den tatsächlichen Sichtweisen der Befragten) und den Forscherhypothesen* hierüber möglich ist und durch ein bestimmtes methodisches Vorgehen (bspw. Rückfragen bei den Interviewten) sichergestellt werden kann. Auch Lincoln und Guba schlagen ein ähnliches Vorgehen unter der Bezeichnung "*member check*" vor "*whereby data, analytic categories, interpretations, and conclusions are tested (!) with members from whom the data were originally collected*" (LINCOLN, GUBA 1985, S.314), und darüber hinausgehend weitere Strategien vor, mit deren Hilfe die "*Glaubwürdigkeit*" (*credibility*) der Ergebnisse sichergestellt werden kann: eine *verlängerte Teilnahme* und *beständige Beobachtung* im Forschungsfeld, *Triangulation* von verschiedenen *Methoden, Datenquellen* und *Untersuchern*, *Supervision* durch andere Forscher und *die Analyse von negativen Fällen*. Welche methodologische Bedeutung die Ersetzung des Begriffs "*interne Validität*" durch den Begriff "*Glaubwürdigkeit*" haben soll, bleibt dabei allerdings unklar. Einerseits kann die Einführung dieses Begriffes als Gütekriterium so verstanden werden, daß der Forscher sich um eine konsistente, verständliche und plausible Darstellung seiner Ergebnisse bemüht. Die Validitätsfrage wäre hiermit tatsächlich überhaupt nicht angesprochen, denn *glaubwürdige, konsistente und verständliche Berichte* über soziale Phänomene können natürlich *falsch* sein, ebenso wie hochgradig unglaubwürdige Berichte reale Phänomene zutreffend beschreiben können. Auf eine solche Deutung des Begriffs Glaubwürdigkeit würde die Kritik von Uta Gerhardt in vollem Umfang zutreffen, die Glaser und Strauss - beide Autoren hatten in ihrer 1967 erschienenen Monographie ebenfalls den Begriff Glaubwürdigkeit als Gütekriterium eingeführt - vorwirft, damit die "*Gültigkeitsfrage (...) auf Probleme der Didaktik*" zu reduzieren (GERHARDT 1985, S.238). Die konkreten Maßnahmen, die Lincoln und Guba zur Erhöhung der Glaubwürdigkeit vorschlagen, beschränken sich allerdings nicht auf eine Beschreibung didaktischer Verfahren. Vielmehr werden eine Reihe von Vorschlägen gemacht, wie Fehlinformationen (die etwa durch unzuverlässige Informanten entstehen) vermieden werden oder Annahmen des Forschers überprüft werden können. Auch das von den beiden Autoren vorgeschlagene Verfahren des "*member checks*" stellt letztendlich eine Überprüfung von Forscherhypothesen dar: dabei wird die Richtigkeit der aus dem Datenmaterial entwickelten Interpretationen durch eine erneute Befragung kontrolliert, ein Vorgehen, daß keinesfalls an eine "*Konsenstheorie der Wahrheit*" oder an eine spezifische qualitative Forschungslogik gebunden ist. Den "member check" empfahl bereits Zetterberg der quantitativen Surveyforschung in seiner 1953 erstmals

erschienenen Monographie über "Theorie und Verifikation in der Soziologie": "*When the interview method is employed to obtain factual information, cross checks can often be made (...) We can come back to the same respondent and ask him again (...)*" (ZETTERBERG 1965, S.123). Schließlich verweisen Lincoln und Guba auf die Methode der "*negativen Fallanalyse*" zur Absicherung der Glaubwürdigkeit qualitativer Forschungsergebnisse. Autoren, die diese Methode angewandt und entwickelt haben und die von Lincoln und Guba ausführlich zitiert werden, wie bspw. Cressey, sprechen explizit von der "*Testung von Hypothesen*": hierbei wird in umfangreichem qualitativem Datenmaterial systematische widersprechende Evidenz zu den Forscherhypothesen gesucht.

Die Forderung nach "Glaubwürdigkeit", die Lincoln und Guba aufstellen, bedeutet also nicht, daß allein eine möglichst konsistente und plausible *Darstellung* der Forschungsergebnisse ausreichend ist. Vielmehr wird damit postuliert, daß der Leser sich darauf verlassen kann, daß bestimmte methodische Regeln im Forschungsprozeß eingehalten wurden: Solche methodischen Regeln umfassen u.a. die Überprüfung von Forscherhypothesen, indem Informanten mehrfach befragt werden oder indem Prognosen über das Eintreten bestimmter Ereignisse aufgestellt werden. Es ist dies allerdings keine andere Glaubwürdigkeit, wie sie auch von statistischen Ergebnissen verlangt wird: auch hier wird normalerweise erwartet, daß alle möglichen methodischen Vorkehrungen getroffen wurden, um mögliche Verzerrungen und Fehler auszuschließen. Es erscheint also kaum angemessen, den Begriff der Glaubwürdigkeit als ein alternatives Gütekriterium für die qualitative Sozialforschung einzuführen, zumal er Anlaß zu dem Mißverständnis gibt, daß für die qualitative Sozialforschung bereits die *plausible Darstellung von Forschungsergebnissen*, also im Zweifelsfall auch die Präsentation "Potemkinscher Dörfer", ausreichend sei.

2.2.2 "Übertragbarkeit" bzw. "Anwendbarkeit" statt "externer Validität"

Lincoln und Guba halten das Konzept der "externen Validität" für ungeeignet für die naturalistische Feldforschung, weil es auf einer *nomothetischen* Grundlage beruhe - der Auffassung, wonach soziale Phänomene ebenso wie Naturereignisse von zeit- und kontextunabhängigen "*rules of nature*" bestimmt würden (LINCOLN, GUBA 1985, S.297). Weil es im Untersuchungsbereich der Sozialwissenschaften jedoch keine deterministischen, kontextfreien Phänomene gebe ("*It is difficult to imagine a human activity that is context-free.*" (ebd. S.114)), sei die Forderung nach Verallgemeinerung

grundsätzlich verfehlt. Lincoln und Guba verweisen dazu auf das bekannte Phänomen, daß sich statistische Schlußfolgerungen mit umso größerer Präzision treffen ließen, je mehr die untersuchte Population in homogene Schichten unterteilt würde. Dieser Effekt würde nur deswegen eintreten, weil die Unterteilung in homogene Schichten zu einer Angleichung von *Kontexten* führe. Die Forderung nach Verallgemeinerbarkeit müsse also in den Sozialwissenschaften ersetzt werden durch die Forderung nach "*Übertragbarkeit*" (*transferability*) und "*Anwendbarkeit*" (*applicability*): die zentrale Frage müßte sein, inwieweit sich die in einem Kontext gefundenen Ergebnisse auf einen anderen übertragen bzw. in einem anderen Kontext anwenden lassen.

Es lohnt sich, auf dieses Kontextargument etwas ausführlicher einzugehen, weil Lincoln und Guba hier an eine klassische Kritik des interpretativen Paradigmas an der quantitativen Soziologie anknüpfen, die sich direkt aus dessen handlungstheoretischen Prämissen ergibt: Mit dem Verweis auf die Bedeutung von Kontexten wurde der Versuch der traditionellen Surveymethodologie angegriffen, die Zuverlässigkeit methodischer Werkzeuge durch eine Standardisierung der verwendeten Begriffe zu sichern (CICOUREL 1974). Hiergegen wurde der Einwand erhoben, daß sprachliche Zeichen wie alle signifikanten sozialen Gesten i.d.R. keine einheitlich "lexikalische", sondern eine mehr oder minder große Anzahl z.T. unscharf umrissener unterschiedlicher Bedeutungen aufweisen, die von den Zeichenbenutzern kontextabhängig ausgewählt und konkretisiert werden müssen. In dem Maße, in dem versucht wird, diese Kontextabhängigkeit durch die Verwendung vermeintlich objektiver Begriffe zu heilen, verliere die Soziologie ihren eigentlichen Gegenstand: die reale Bedeutung des jeweiligen signifikanten Symbols in seinem konkreten Kontext und damit die des jeweiligen sozialen Handelns in spezifischen Situationen gehe verloren (vgl. WEINGARTEN, SACK 1976, S.17).

In welcher Beziehung steht dieses Kontextargument zum Problem der *allgemeinen Geltung* von sozialen Regeln und Normen und damit zum methodologischen Problem der *Verallgemeinerung*? Es impliziert nicht notwendigerweise, daß die Festlegung der Bedeutung sozialer Symbole in jeder einzelnen Situation vollständig neu erfolgt. Vielmehr ist Konzepten wie "Lebenswelt" oder "soziale Lebensform", die einen festen Platz im interpretativen Paradigma einnehmen, die Vorstellung inhärent, daß Gesellschaftsmitglieder in ihrem Alltagshandeln auf gemeinsam geteilte *Wissensbestände* zurückgreifen, um Kontexte zu identifizieren, d.h. um Situationsmerkmale ausfindig zu machen, die ihnen einen Hinweis darauf geben, *welche*

Symbole und signifikante Gesten *in welcher Weise* zu verstehen sind. Die Existenz dieser Wissensbestände und der in ihnen enthaltenen *Regeln* für (1.) die Identifikation von Kontexten und (2.) einen angemessenen Gebrauch und das Verstehen von signifikanten Gesten und Symbolen in diesen Kontexten sind unverzichtbare Grundlagen für soziale Interaktion überhaupt. Wenn soziale Akteure nämlich nicht bis zu einem gewissen Umfang in der Lage sind, Kontexte einheitlich wahrzunehmen und bestimmte Gesten in diesen Kontexten einheitlich zu handhaben, oder bei ihren Interaktionspartnern diese Fähigkeiten nicht erwarten können, wäre soziale Ordnung schlechterdings nicht möglich¹. Jene Regeln, auf die hierbei (zumeist routinemäßig und stillschweigend) zurückgegriffen wird, können selbst wiederum an bestimmte kulturelle oder sozioökonomische Kontexte gebunden sein und dann natürlich nicht als "*zeit- und kontextunabhängige Naturgesetze*" betrachtet werden. Das bedeutet aber nicht, daß sie keinen *allgemeinen Charakter* tragen würden. Die Angehörigen einer bestimmten Kultur oder Lebensform müssen selber in der Lage sein (etwa bestimmte Situationsmerkmale) zu *verallgemeinern*, um zu wissen, in welchen Kontexten sie sich befinden, welches Verhalten man hier von ihnen typischerweise erwartet und welches sie von ihren Interaktionspartnern erwarten dürfen. Der Untersucher sozialer Lebensformen oder Kulturen steht vor einer ähnlichen Aufgabe: er muß sich ein Wissen darüber erwerben, welche Situationsmerkmale in der untersuchten Kultur als allgemeine Kennzeichen für einen bestimmten Kontext betrachtet werden und die für diesen Kontext geltenden Regeln erfassen. Anderenfalls ist er nicht in der Lage, die für eine Kultur *typischen* sozialen Phänomene (bestimmte Handlungsweisen, Praktiken, die Bedeutung bestimmter Symbole etc.) von zufälligen Erscheinungen zu unterscheiden. Er ist deshalb darauf angewiesen, bestimmte beobachtete Phänomene zu verallgemeinern, auch dann, wenn er gar nicht den Anspruch formuliert, generelle Gesetzmäßigkeiten sozialen Lebens zu finden, die unabhängig von raum-zeitlichen historischen Bezügen existieren.

"Verallgemeinerung" kann also nicht als *Alternative* zur "*Übertragung*" oder "*Anwendung*" von Untersuchungsergebnissen auf andere Kontexte betrachtet werden, wie dies Lincoln und Guba tun; ein Akt der Verallgemeinerung ist vielmehr die *Voraussetzung* für die erfolgreiche Übertragung von Untersuchungsergebnissen. Folgt

¹ Gute Beispiele für das Zusammenbrechen sozialer Ordnungen auf der Mikroebene, wenn diese Fähigkeiten bewußt außer Kraft gesetzt werden, liefern die Krisenexperimente von Garfinkel (vgl. WEINGARTEN, SACK, SCHENKEIN 1976, S.49).

man nämlich der Argumentation von Lincoln und Guba, so sind Forschungsergebnisse dann auf einen anderen Untersuchungsgegenstand übertragbar, wenn sich die jeweiligen Kontexte *gleichen* oder zumindest *ähneln*. Eine begründete Entscheidung darüber, ob sich zwei Kontexte gleichen oder ähneln, setzt allerdings voraus, daß es möglich ist, jene *allgemeinen Merkmale* zu identifizieren, die für die betreffenden Kontexte essentiell sind. In diesem Sinne bedeutet auch der "klassische" bzw. "konventionelle" Gebrauch des Begriffs der externen Validität nichts anderes als Übertragbarkeit. In der experimentell und quantitativ orientierten Sozialforschung bezieht sich externe Validität auf "*the extent to which one can generalize across various kinds of persons, settings, times (...)*" (COOK, CAMPBELL 1979, S.73). Die Frage danach, inwieweit die Ergebnisse einer Studie über die jeweils verwendeten Untersuchungsgruppen und experimentellen Settings hinaus *verallgemeinert* werden können, ist gleichbedeutend mit der Frage, inwieweit sie auf andere Gruppen und Settings *übertragen* werden können (vgl. SMITH, GLASS 1987, S.6). Bei der Bewertung qualitativer Feldstudien läßt sich eine ähnlich bedeutsame Frage formulieren: inwieweit können die von einem Untersucher in einem bestimmten Beobachtungssetting gesammelten Daten über eine Kultur oder soziale Lebensform und die hieraus gewonnenen Ergebnisse verallgemeinert werden, d.h. inwieweit sind sie übertragbar auf andere Situationen (oder inwieweit könnten sie von anderen Beobachtern auch gewonnen werden):

Können die Äußerungen bestimmter junger Frauen in einem bestimmten Dorf in Samoa gegenüber der Forscherin Margaret Mead verallgemeinert werden auf das Sexualverhalten junger Samoaner oder handelt es sich um Artefakte der Befragungssituation, etwa um unzutreffende Berichte unzuverlässiger Informanten? Geben diese Berichte, bedingt durch eine (möglicherweise zufällig) verzerrte Auswahl der Gesprächspartner, nur das Verhalten einer Minderheit wieder? Dasselbe gilt für die beiden anderen, bereits genannten Beispiele: Geben die mündlichen Berichte der Sozialhilfeempfänger oder die Akten des Sozialamtes bestimmte objektive Daten über den Lebenslauf zutreffend wieder? Reflektieren die von Kirk und Miller erhobenen Informationen über den Kokakonsum in Peru tatsächlich ausgeübte kollektive Praktiken in diesem Land oder sind diese Informationen verzerrt, weil die Informanten Sanktionen bei der Preis-gabe von Praktiken des Drogenkonsums fürchten?

Hiermit sind Fragen der *externen Validität* angesprochen, wenn man die Definition von Cook und Campbell zugrundelegt, wonach sich externe Validität auf das Ausmaß

bezieht, in dem Forschungsergebnisse über das jeweilige Untersuchungssetting und die jeweilige Untersuchungsgruppe hinaus verallgemeinert werden können.

2.2.3 Das Risiko der "rhetorischen Validierung"

Die Diskussion der von Lincoln und Guba vorgenommenen Gegenüberstellung zwischen den Begriffen "Glaubwürdigkeit" und "interne Validität" einerseits und "Übertragbarkeit" bzw. "Anwendbarkeit" und "externe Validität" andererseits zeigt, daß die Ersetzung traditioneller Gütekriterien durch neuartige Begriffe methodologisch keinen Gewinn bringt, dafür aber mit nicht zu unterschätzenden Risiken verbunden ist:

- * Entweder wird das Gültigkeitsproblem auf diese Weise auf ein Problem der rhetorisch und didaktisch geschickten Präsentation von Forschungsergebnissen reduziert,
- * oder es werden Gütekriterien formuliert und Wege der Geltungssicherung vorgeschlagen, die sich in ihren Grundprinzipien von den klassischen Standards kaum unterscheiden.

Die Überlegungen von Lincoln und Guba bewegen sich zwischen diesen beiden Extremen hin und her. Bei der Gegenüberstellung der Begriffe "Glaubwürdigkeit" und "interner Validität" führen sie eine neuartige Terminologie ein, um letztendlich relativ konventionelle Wege der Validitätssicherung - etwa die Prüfung der Forscherhypothesen durch wiederholte Beobachtung und Befragung, die systematische Suche nach widersprechender Evidenz etc. - zu empfehlen. Hinsichtlich der Bedeutung von Generalisierbarkeit führen sie, während sie die vermeintlich "naiv realistischen" konventionellen Vorstellungen von Repräsentativität und Generalisierbarkeit kritisieren, an anderer Stelle das Konzept der "*holistischen Generalisierung*" ein, das in seiner Konsequenz zu einem strengen Begriff von Repräsentativität führt: Holistische Generalisierung ist dann möglich, wenn - wie in einem Hologramm - ein Teil eines untersuchten Phänomens alle Eigenschaften des gesamten Phänomens enthält "*that full information about a whole is stored in its parts, if only we knew how to get at it*" (LINCOLN, GUBA 1985, S.127). Eine Untersuchungsgruppe, die "holistische Generalisierungen" ermöglichen kann, stellt ein Teil des "Ganzen" dar, welches *alle* Informationen über dieses Ganze enthält: "(...) *any part or component is a perfect sample in the sense that it contains all of the information about the whole that one*

might ever hope to obtain (...) imperfect (blurred) information from any source can be improved (clarified), if one has the appropriate filters for so doing" (ebd., S.128). Die beiden Autoren vermögen allerdings keinerlei Strategien zu nennen, mit deren Hilfe solche "perfekten Samples" gezogen werden können oder jene "angemessenen Filter" konstruiert werden können, die "imperfekte Informationen" verbessern oder klären könnten. Damit gerät das Konzept der "holistischen Generalisierung" in Gefahr, auf eine legitimatorische und rhetorische Funktion beschränkt zu werden: Auf der einen Seite kann hiermit methodologische Kritik an der Zusammensetzung der Untersuchungsgruppe mit dem Hinweis abgewehrt werden, daß Kriterien wie Repräsentativität in der Feldforschung gänzlich ungeeignet seien. Auf der anderen Seite transportiert dieses Konzept einen weitgehenden Wahrheitsanspruch: auf eine ungeklärte und nicht beschreibbare Weise soll sichergestellt sein, daß die Untersuchungsgruppe und die Untersuchungssituation alle Informationen über das untersuchte Phänomen enthalten.

2.3 Müssen die klassischen Gütekriterien den Anforderungen der qualitativen Forschung *angepaßt* werden?

Will man für die qualitative Sozialforschung nicht auf methodische Standards und Gütekriterien vollends verzichten, oder neue Begriffe wie "*Glaubwürdigkeit*" und "*Übertragbarkeit*" einführen, die das Risiko mit sich bringen, daß das Gültigkeitsproblem auf ein Problem der Darstellung von Forschungsergebnissen reduziert wird, so stellt sich erneut die Frage nach der Anwendbarkeit von Begriffen wie externe und interne Validität im interpretativen Paradigma: Welche Bedeutung besitzen solche Konzepte im Kontext einer Forschungstradition, die strikt experimentelle Vorgehensweisen vermeidet und auf eine Standardisierung von Forschungsinstrumenten verzichtet?

Bisherige Versuche, die Konzepte interne und externe Validität im Rahmen interpretativer Forschung zu klären und für deren Bedürfnisse zu reformulieren, haben allerdings sehr uneinheitliche Resultate erbracht:

Volmerg (1983) setzt **interne Validität** mit Reliabilität (Zuverlässigkeit) und **externe Validität** mit "Realitätshaltigkeit" gleich. Wenn, wie in der quantitativen Forschung, allein die Suche nach Reliabilität im Mittelpunkt stehe, so sei ein Verlust von Alltagsnähe und Realitätshaltigkeit, d.h. von externer Validität die Folge. Da der Alltagsnähe und der Realitätshaltigkeit aus der Sicht des interpretativen Paradigmas eine besondere Priorität zukomme, müsse das Hauptaugenmerk des Untersuchers der alltagsnahen Gestaltung der Erhebungssituation gelten. Die entscheidende Frage dabei sei: *"Wie haben die Erforschten die Situation, in der die Erhebung stattfand, wahrgenommen und für sich definiert. Konnte die Situation als eine identifiziert werden, in der man sich zu dem Forschungsgegenstand so verhält wie unter natürlichen Bedingungen?"* (VOLMERG 1983, S.128). Hiermit ist für Volmerg das Problem der internen Validität ebenfalls gelöst, denn *"wie zuverlässig der Realitätsgehalt der Erhebungssituation ist, entscheidet dann über die interne Gültigkeit der Methoden"* (ebd., S.129).

Im Unterschied zu Volmerg unterscheiden LeCompte und Preissle (1993) zwischen interner und externer Validität einerseits und Reliabilität andererseits. Das Konzept der **internen Validität** *"raises the problem of whether conceptual categories understood to have mutual meanings between participants and the observer are shared"* (LECOMPTE, PREISSLE 1993, S.342). Weiterhin seien damit folgende Probleme angesprochen:

- * das Problem der Reifung und geschichtlicher Einflüsse - der Forscher müsse in der Lage sein, Reifungsprozesse von anderen intervenierenden Einflüssen zu unterscheiden,
- * Beobachtereinflüsse, die sich aus der spezifischen Rolle des Untersuchers im Feld ergeben,
- * Probleme, die sich aus der selektiven Sammlung von Daten vor allem dann ergeben, wenn der Gegenstandsbereich sehr komplex ist,
- * "Mortalität", d.h. Veränderungen in der Untersuchungsgruppe während des Untersuchungszeitraums
- * und die Gefahr von fehlerhaften Schlußfolgerungen.

LeComptes und Preissles Definition von **Externer Validität** setzt sich aus zwei Teilen zusammen. Der erste, genannt "*Vergleichbarkeit*" (*comparability*) ähnelt dem von Lincoln und Guba vorgeschlagenen Konzept der "Übertragbarkeit": "(...) *comparability* (...) is the degree to which the components of a study, including the units of analysis, concepts generated, population characteristics, and setting, are sufficiently well described and defined that other researchers can use the results to compare to other studies addressing related issues" (ebd., S.348). Der zweite Bestandteil externer Validität ist LeCompte und Preissle zufolge "*Übersetzbarkeit*" (*translatability*) "*the degree to which the researcher uses theoretical frames, definitions and research techniques accessible to and understood by other researchers in the same or related disciplines.*" (ebd., S.348, 349)

Unter **Reliabilität** verstehen die Autorinnen die Replizierbarkeit von Untersuchungen und unterscheiden dabei die Begriffe *interne* und *externe Reliabilität*. Externe Reliabilität bezieht sich auf die Wiederholbarkeit einer bestimmten Studie und interne Reliabilität soll das Ausmaß bezeichnen, in dem verschiedene Untersucher in derselben Studie zu denselben Ergebnissen gelangen.

Insgesamt läßt sich festhalten, daß bislang keine einheitliche Sprachregelung für die Übertragung der klassischen Gütekriterien auf die qualitative Sozialforschung gefunden wurde. Auch wenn die Begriffe Reliabilität, interne Validität und externe Validität

gebraucht werden, wird ihre genaue Bedeutung in der Regel nicht expliziert, sie werden von verschiedenen Autoren uneinheitlich und teilweise inkonsistent² verwendet. Dieser Umstand läßt sich u.a. darauf zurückführen, daß eine Auseinandersetzung mit klassischen Definitionen der Gütekriterien und eine systematische *Übertragung* auf die qualitative Sozialforschung i.d.R. nur in begrenztem Umfang versucht wird. Volmerg etwa handelt die Bedeutung der Begriffe interne und externe Validität im Kontext experimenteller Forschung und quantitativer Methodologie nur kursorisch ab und stellt dann eine eigene Definition vor, die von dem klassischen Verständnis weitgehend abweicht³, ohne diese Abweichung (etwa durch die besonderen Anforderungen interpretativer Sozialforschung) zu begründen. LeCompte und Preissle votieren, weil sich die klassischen Gütekriterien auf qualitative Feldforschung nicht direkt übertragen ließen, für deren *metaphorischen Gebrauch*, ohne allerdings zu verdeutlichen, worin ein solcher metaphorischer Gebrauch bestehen kann.⁴ (LECOMPTE, PREISSLE 1993, S.332)

Da sich weder der Verzicht auf methodische Standards noch die Entwicklung alternativer Gütekriterien als ein gangbarer Weg erwiesen haben, erscheint eine Klärung des Validitätsbegriffs im Kontext der qualitativen Sozialforschung dringend erforderlich. Sollen die konventionellen Gütekriterien für deren Anforderungen "angepaßt" bzw. "reformuliert" werden, so muß jedoch zuerst deren Bedeutung in ihrem klassischen Kontext, der Methodik standardisierter Befragungen und experimenteller Forschung, herausgearbeitet werden. Eine Übertragung von Validitätskonzepten auf eine Forschungstradition, die sich in ihrem methodologischen Selbstverständnis und in ihren Methoden der Designgestaltung, der Datenerhebung und -auswertung radikal von

² So fassen etwa LeCompte und Preissle unter dem Konzept der internen Validität, das sie anfänglich mit einem homogenen Sprachgebrauch von Forschern und Beforschten gleichsetzen, anschließend auch solche Phänomene, die ihrer eigenen Definition gemäß eher dem Bereich der externen Validität zuzuordnen sind, wie etwa Probleme, die sich aus der selektiven Sammlung von Daten ergeben.

³ Insbesondere, indem sie "interne Validität" und "Zuverlässigkeit" einfach gleichsetzt.

⁴ Tatsächlich verwenden sie dann einige der Begriffe zur Kennzeichnung von Gütekriterien nicht metaphorisch, sondern wortwörtlich, so etwa den Begriff Reliabilität: "*Reliability refers to the extent to which studies can be replicated.*" (LeCompte, Preissle 1993) Eine Fassung der traditionellen Definition lautet etwa: "*Reliabilität bezeichnet die Reproduzierbarkeit von Ergebnissen unter den gleichen intersubjektiven Bedingungen, also insbesondere die Forderung, daß anderen Forscher bei Anwendung desselben Erhebungs-instruments in Interaktion mit demselben Untersuchungsgegenstand zu demselben Ergebnis gelangen.*" (KRIZ, LISCH 1988, S.219 f.). LeCompte und Preissle räumen ein, daß in der qualitativen Forschung wegen der Natur des Datenmaterials und des Forschungsprozesses dieser Forderung nur schwer entsprochen werden könnte, machen aber nicht deutlich, wie das Kriterium der Zuverlässigkeit stattdessen gefaßt werden müßte, um in der qualitativen Sozialforschung anwendbar zu sein.

hypothetiko-deduktiver Sozialforschung abhebt, läßt sich allerdings nur dann sinnvoll vornehmen, wenn eine Unterscheidung getroffen wird zwischen dem *erkenntnistheoretischen Kern* von Validitätskonzepten einerseits und deren *operationalen Ausgestaltungen* andererseits.

- * Nur der erkenntnistheoretische Kern von Begriffen wie interne Validität und externe Validität, Reliabilität und Repräsentativität kann sich auf Fragen beziehen, die in der empirischen Sozialforschung unabhängig von der methodologischen Orientierung des jeweiligen Untersuchers eine Rolle spielen. Es sind dies Fragen wie: *Lassen sich die Untersuchungsergebnisse auf andere Settings und Untersuchungsgruppen verallgemeinern bzw. übertragen? Würden andere Untersucher zu den gleichen oder weitgehend ähnlichen Ergebnissen gelangen? Wurden alternative Erklärungen geprüft?* etc.
- * Validierungsstrategien stellen die operationale Ausgestaltung von Validitätskonzepten im *Kontext einer bestimmten Forschungstradition*⁵ dar. Validierungsstrategien wie sie die Überprüfung der Split-half- und Paralleltestreliabilität, der Randomisierung von Versuchs- und Kontrollgruppen oder der statistischen Korrelation zwischen Testerfolg und einem Außenkriterium darstellen, sind nur im Rahmen eines bestimmten (testpsychologischen, experimentellen oder quasi-experimentell statistischen) Forschungsdesign sinnvoll einsetzbar. Im Kontext einer qualitativen Forschungstradition, die keine standardisierten Meßinstrumente verwendet und in ihren Untersuchungsdesigns nicht in ähnlicher Weise statistische oder experimentelle Kontrolle einbauen kann, müssen Validitätskonzepte anders operationalisiert werden. Oder anders gesprochen: die qualitative Sozialforschung benötigt nicht *alternative Gütekriterien*, sondern *alternative Validierungsstrategien*.

Im Folgenden sollen verschiedene klassische Validitätskonzepte dargestellt und diskutiert werden. Im Anschluß daran wird versucht, den erkenntnistheoretischen Kern der Begriffe *interne* und *externe Validität* herauszuarbeiten und verschiedene Validierungsstrategien aus dem Kontext qualitativer Sozialforschung (*Kommunikative Validierung*, *Diskursive Validierung*, *Exhaustation von Deutungshypothesen*,

⁵ Zum Begriff der "Forschungstradition" vgl. LAUDAN 1977.

Analytische Induktion, Verfahren qualitativer Stichprobenziehung, Triangu-lation)
hierauf zu beziehen.

3 Klassische Validitätskonzepte der quantitativen Sozialforschung

Untersucht man gängige Definitionen des Validitätsbegriffs (etwa in Handbüchern der Methodenlehre), so stellt man fest, daß epistemologischer Kern und operationale Ausgestaltungen oft kaum voneinander zu trennen sind: die Autoren legen solche Gütekriterien i.d.R. mit Bezug auf die benutzten Meßinstrumente und -verfahren fest. Hierbei werden methodologische Aspekte oft nur sehr global angesprochen⁶. Im Folgenden soll deshalb auf mehrere klassische Aufsätze zum Validitätsbegriff in der quantitativen Sozialforschung zurückgegriffen werden: auf die Arbeiten von Louis Guttman, Hans Zetterberg sowie Thomas D. Cook und Donald T. Campbell.

3.1 Guttman

Guttman, der in einem 1950 erschienenen Aufsatz methodische und methodologische Probleme der *Skalierung in der Einstellungsforschung* behandelt (GUTTMAN 1950, S. 57ff), unterscheidet zwischen *interner* und *externer Validität*.

Interne Validität betrifft demnach die Beziehung zwischen dem Inhalt von Items eines Fragebogens und dem "Universum", das gemessen werden soll. Es stelle sich für jedes Item die Frage, ob es zum untersuchten Universum gehöre, d.h. ob die Antwort auf diese bestimmte Frage zur Erforschung dieses Universums beitragen könne. Eine Entscheidung über diese Frage könne - so Guttman - nur durch eine *inhaltliche* Beurteilung des Items getroffen werden: *"Only a judgement of the content can answer this, according to our approach. For example, does the following item belong to a universe of knowledge of arithmetic: 'How much is two plus two?' (...) If one wishes to study knowledge of arithmetic, then one asks arithmetic questions. (...) What questions go into each of such universes is determined by the content involved."* (GUTTMAN 1950, S. 57).

In bezug auf die **externe Validität** bestimmt Guttman das *Problem der Prognose* als zentrale Fragestellung. Während jedes Universum nur eine interne Validität besäße,

⁶ Vergleiche etwa die Definition von Schnell und Kollegen (1989, S. 150), wonach die Validität eines Meßinstruments *"das Ausmaß, in dem das Meßinstrument tatsächlich das mißt, was es messen soll"* darstellt.

habe es viele mögliche externe Validitäten, da auf der Basis eines Index eine Vielzahl von Prognosen möglich seien: *"Knowledge of arithmetic may have some validity for predicting success as an accountant, a different validity for predicting success as an engineer, etc, etc"* (GUTTMAN 1950, S. 57). Voraussetzung zur Bestimmung der externen Validität eines Index ist demnach die genaue *Definition* des vorherzusagenden Phänomens oder Ereignisses und die Bestimmung der *internen Validität* des zur Vorhersage benutzten Index: *"'Internal validity' must be agreed on first."* (GUTTMAN 1950, S. 58). Erst danach könne die Korrelation zwischen Prognoseinstrument und definiertem Kriterium bestimmt werden. Die externe Validität eines Index ist für Guttman dabei nicht zu verwechseln mit der Realitätsübereinstimmung einer Aussage, die ein Befragter macht; auch "un-wahre" oder "ungenau" Aussagen von Befragten lassen möglicherweise Vorher-sagen über bestimmte Phänomene zu⁷: *"A man's statement about his age might not be accurate chronologically, but it might prove to be a good predictor, say, of his personal adjustment or other areas defined exteranally to it."* (GUTTMAN 1950, S. 59).

3.2 Zetterberg

Auch Zetterberg unterscheidet zwischen interner und externer Validität (ZETTER-BERG 1965). Interne Validität betrifft das Verhältnis zwischen Theorie und Daten: *"(it) is the extent to which an indicator corresponds to a definition"* (ZETTERBERG 1965, S. 14). Während externe Validität sich auf empirische Verhältnisse bezieht, drückt also interne Validität eine rein logische Beziehung aus: *"Internal validity, in other words, can be appreciated without empirical studies, while the determina-tion of external validity is a test of a hypothesis."* (ZETTERBERG 1965, S. 115)

Perfekte interne Validität bedeutet nach Zetterberg *"that the indicator has the same scope of content as the definition"* (ZETTERBERG 1965, S. 115). Ausgehend von dieser Definition bestimmt Zetterberg am Beispiel eines Index für "Arbeitszu-friedenheit" drei Probleme für die interne Validität:

⁷ Was Guttman hier anspricht, könnte in Anlehnung an Gerhardt 1985 die "ad quem-Struk-tur" von quantitativen Daten genannt werden.

- * Der benutzte Indikator mißt zwar den Gegenstand, deckt ihn aber bei weitem nicht ab.
- * Der benutzte Indikator mißt den Gegenstand, zusätzlich aber auch noch etwa, das nicht zum Gegenstand gehört.
- * Der benutzte Indikator und der Gegenstand haben zwar eine Schnittmenge, zum Meßbereich des Indikators gehören jedoch zusätzliche Phänomene, die nicht zum Gegenstandsbereich gehören (und *vice versa*: der Indikator erfaßt bei weitem nicht alle Phänomene, die dem Gegenstandsbereich angehören).

Interne Validität sei - so Zetterberg - dadurch abzusichern, daß die Effekte der o.g. Fehler minimiert werden, indem bspw. mehrere Indikatoren zu einem Index kombiniert werden, wobei sich die *Fehlerkomponenten der einzelnen Indikatoren* durch die stärkere Verteilung gegenseitig aufheben. (ZETTERBERG 1965, S. 118)

Demgegenüber sei externe Validität (ebd., S. 120 ff) dann bedeutsam, wenn ein Index oder Indikator zur Prognose eines anderen Indikators genutzt werden solle, bspw. die Angaben über zukünftige Handlungsabsichten einer bestimmten Gruppe von Befragten als Prognoseinstrument für ihr reales zukünftiges Verhalten. Diese Angaben oder Aussagen seien aber ggf. durch verschiedene Faktoren verfälscht: Befragte könnten bspw. lügen bzw. ihre Antworten in Richtung auf sozial erwünschte Äußerungen modifizieren. Solche *Verfälschungen* oder *Verzerrungen* seien prinzipiell durch Prüfungen der Datenkonsistenz oder wiederholte Befragungen aufspürbar. Diese Tests seien jedoch nicht in der Lage, die externe Validität von Daten sicherzustellen.

3.3 Cook und Campbell

Zusätzlich zu den Begriffen interne und externe Validität führen Cook und Campbell (1979) für den Bereich der *experimentellen* bzw. quasi-experimentellen Forschung die Konzepte "*statistical conclusion validity*" und "*construct validity*" ein.

- * Die **statistical conclusion validity** betrifft bspw. die Frage, ob der benutzte Test sensitiv genug ist (Problem der Teststärke bzw. des Fehlers zweiter Ordnung) sowie die Frage, ob die Effektstärke ausreichend ist (vgl. COOK, CAMPBELL 1979, S. 39) - Fragen, die sich auf die verwendeten statistischen Teststrategien beziehen.
- * Demgegenüber fassen Cook und Campbell unter dem Begriff der **internen Validität** Fragen, die die Kausalität der angenommenen Beziehung zwischen Variablen betreffen, da statistische Korrelation nicht zwangsläufig mit kausaler Beziehung gleichgesetzt werden kann: "*Once it has been established that two variables covary, the problem is to decide whether there is any causal relationship between the two and, if there is, to decide whether the direction of causality is from the measured or manipulated A to the measured B, or vice versa.*" (ebd., S. 50)
- * Hiervon grenzen Cook und Campbell die **construct validity** ab, einen Begriff, den sie von Cronbach und Meehl (1955) übernehmen. Construct validity hängt mit der Gültigkeit der zugrundeliegenden theoretischen Konstrukte zusammen. Eine Kausalbeziehung zwischen einem Merkmal A und einem Merkmal B ($A \Rightarrow B$) lässt sich in einem anderen theoretischen Modell ggf. durch eine Beziehung zwischen X und B ($X \Rightarrow B$) oder sogar zwischen X und A einerseits und X und B andererseits ($X \Rightarrow A \wedge X \Rightarrow B$) erklären. Die construct validity betrifft damit die Frage, inwieweit alternative Erklärungsansätze zum behaupteten theoretischen Konstrukt auf der Basis des empirischen Materials plausibel sind. Hiermit wird also - ebenso wie mit dem Begriff der internen Validität - die Frage nach intervenierenden Variablen in die Validitätsdiskussion einbezogen⁸.

⁸ Der Begriff der internen Validität bezieht sich bei Cook und Campbell eher auf die Gestaltung der experimentellen Situation (die die Möglichkeit intervenierender Einflüsse möglichst ausschließen soll), der Begriff der Konstruktvalidität eher auf die Gefahr, ein experimentelles Ergebnis mit

- * Unter Fragen der **externen Validität** fassen Cook und Campbell schließlich Problembereiche, die sich aus der Verallgemeinerung von Ergebnissen der Untersuchung von Stichproben oder Experimentalgruppen auf (Teil-) Populationen ergeben: Diese haben nach Cook und Campbell zu tun mit "(1) *generalizing to particular target persons, settings, and times*, and (2) *generalizing across types of persons, settings, and times*" (ebd., S. 71).

Für alle vier Bereiche arbeiten Cook und Campbell die ihnen bedeutsam erscheinenden **Validitätsbedrohungen** ("*threats to validity*") heraus - etwa geringe Test-stärke oder Voraussetzungsverletzungen bei statistischen Tests für die statistical construct validity (ebd., S. 42), Interaktionen unterschiedlicher "Treatments" für die construct validity (ebd., S. 67 f) oder Interaktionen zwischen Sampleauswahl und Treatment für die externe Validität (ebd., S. 73). Sodann entwickeln sie z.T. Vorschläge, wie diesen konkreten Validitätsbedrohungen begegnet werden kann: Hierzu gehört etwa die theoriegeleitete Stichprobenauswahl zur Erhöhung der externen Validität (ebd., S. 79 f), oder verschiedene Randomisierungstechniken zur Erhöhung der internen Validität (ebd., S. 55 f)

einem unzureichenden theoretischen Konstrukt zu erklären. Diese Unterscheidung besitzt also vor allem Sinn für die experimentelle Forschung.

3.4 Der "*fallibilistische Kern*" klassischer Validitätskonzepte

Guttman, Zetterberg sowie Cook und Campbell konzeptualisieren Validität für eine spezifische Forschungsmethodik: Guttman nimmt bezug auf die Einstellungsuntersuchungen während des Zweiten Weltkriegs, an denen er beteiligt war, Zetterberg bezieht sich u.a. auf Ergebnisse der Wahl- und Evaluationsforschung, Cook und Campbell auf die psychologische experimentelle oder quasi-experimentelle Forschung. Die von diesen Autoren entwickelten Konzepte besitzen jedoch eine Gemeinsamkeit: *Validitätsbedrohungen*, *Irrtumsmöglichkeiten* und *Fehlerquellen* stehen im Vordergrund. So bestimmt etwa Zetterberg typische logische Fehler, die bei der Konstruktion eines Index gemacht werden können, als Probleme interner Validität. Bei der Diskussion des Begriffs externe Validität konzentriert er sich auf *Verfälschungen* und *Verzerrungen*, die von der Untersuchungssituation ausgehen. Maßnahmen zur Erhöhung der externen Validität könnten die Existenz solcher Verzerrungen aufdecken, nicht jedoch die Gültigkeit oder Wahrheit der Ergebnisse im strengen Sinne belegen: "*Such tests of agreement and consistency can never fully prove validity, but they can fully disprove validity.*" (ZETTERBERG 1965, S. 123) Cooks und Campbells Ziel ist es schließlich, eine (möglichst vollständige) Taxonomie von Fehlerquellen ("*threats for validity*") nebst jenen Strategien zu entwickeln, die zu deren Aufdeckung und zur Abwehr von Validitätsbedrohungen dienen können.

Die Validität eines Forschungsergebnisses wird innerhalb solcher Konzeptionen nicht als eine perfekte Übereinstimmung zwischen Realität und Forschungsergebnis betrachtet, sondern als Versuch der Annäherung: "*We shall use the concepts validity and invalidity to refer to the best available approximation to the truth or falsity of propositions, including propositions about cause (...) At best, one can know what has not yet been ruled out as false.*" (COOK, CAMPBELL 1979, S. 37) Damit wird der Begriff der Validität *methodisiert*: Ein Ergebnis gilt nicht deswegen als validiert, weil eine Übereinstimmung mit der Realität festgestellt würde⁹, sondern weil bestimmte Vorichtsmaßnahmen getroffen wurden, um Validitätsbedrohungen entgegen zu wirken.

⁹ Eine solche Definition würde im übrigen unweigerlich in den infiniten Regreß münden - denn auf welche Weise kann sichergestellt werden, daß die Realität tatsächlich so beschaffen ist, wie das Untersuchungsergebnis nahelegt.

Die dargestellten Validitätskonzepte besitzen also gewissermaßen einen *fallibilistischen Kern*. Ausgangspunkt bilden Irrtumsmöglichkeiten menschlicher Forschungstätigkeit, Validierungsstrategien werden als Methoden verstanden, um solche Irrtumsmöglichkeiten und Fehlerquellen zu erkennen, ihnen vorzubeugen und Forschungsergebnisse dagegen abzusichern. Dies impliziert ein Modell sozialwissenschaftlicher Forschungstätigkeit, wonach Methoden der Sozialforschung nicht maßstabsgetreue "Abbilder" sozialer Realität liefern, sondern methodenspezifische Artefakte, "Daten" genannt, die als Grundlage für "Modelle" der Realität dienen können. Hinsichtlich der Abbildung ihres Gegenstandsbereichs sind diese Modelle jedoch unzureichend und defizitär - *"Die Daten schieben sich wie eine Wischblende über das, was sie zu repräsentieren vorgeben; das Ergebnis ist ein stark verschwommenes Bild, auf dem die Konturen des Objekts und die Wirkung des Filters nicht mehr auseinanderzuhalten ist."* (BERGMANN 1985, S. 307) Will man diese optische Metapher fortführen, so beziehen sich Überlegungen zur Validität auf die Beschaffenheit dieser Blende bzw. dieses Filters, Validierungsstrategien dienen zu ihrer Korrektur bzw. Aufhellung.

4 Methodenprobleme und Validierungsstrategien qualitativer Sozialforschung

Das oben beschriebene fallibilistische Kernkonzept von Validität läßt sich auch auf die qualitative Sozialforschung übertragen, ohne daß die Notwendigkeit besteht, Gütekriterien und Validierungsstrategien aus dem Kontext quantitativer Sozialforschung zu verwenden, welche den spezifischen Erfordernissen des interpretativen Paradigmas nicht gerecht werden. Statt der qualitativen Forschung unangemessene Standards zu verordnen, kann vielmehr der Weg von Cook und Campbell eingeschlagen werden: ebenso, wie diese Autoren methodenspezifische *threats for validity* für experimentelle und quasi-experimentelle Forschungsdesigns bestimmt haben, müssen jene Methodenprobleme identifiziert werden, die im Kontext qualitativer Forschungsdesigns typischerweise auftreten - Methodenprobleme, die die klassischen Fehler und "Unschärfen" einer qualitativen "Linse" oder "Brille" repräsentieren. Oder anders ausgedrückt: es müssen *Fehlertheorien* für die qualitative Forschung gesucht bzw. entwickelt werden.

Im Folgenden wollen wir, aufbauend auf Erfahrungen verschiedener empirischer Forschungsprojekte des Sonderforschungsbereichs 186, die mit qualitativen Methoden arbeiten, verschiedene Bedrohungen der Validität spezifizieren und Strategien zu deren Abwehr, also Verfahren der Geltungssicherung, diskutieren, die in diesen Projekten eingesetzt werden. Ziel soll dabei nicht etwa eine vollständige und umfassende Taxonomie von Methodenproblemen und Validierungsstrategien qualitativer Sozialforschung sein; vielmehr geht es uns darum, einige im Kontext der am Sfb betriebenen Biographie- und qualitativen Lebenslaufforschung häufig auftretende, möglicherweise "typische" Probleme zu benennen und Lösungen hierfür zu suchen. Hierbei soll die Unterscheidung zwischen *Bedrohungen der internen Validität* und *Bedrohungen der externen Validität*, die von Cook und Campbell getroffen wurde, beibehalten werden. Allerdings werden die Begriffe *interne und externe Validität* dabei auf die spezifische Forschungslogik qualitativer Methoden bezogen.

4.1 Verfahren zur Sicherung der internen Validität

4.1.1 Bedrohungen der internen Validität in der qualitativen Sozialforschung

Das Konzept der internen Validität bezog sich bei Guttman und Zetterberg auf die *Übereinstimmung zwischen den verwendeten Items* und dem *untersuchten Gegenstandsbereich*, während Cook und Campbell sich mit ihrer Definition auf die Gefahr konzentrierten, daß aus dem Datenmaterial logisch unzulässige Schlußfolgerungen über kausale Zusammenhänge gezogen werden. Interne Validität betrifft in beiden Fällen Fragen der *Theoriekonstruktion* und des Verhältnisses zwischen *Daten und Theorie*: Inwieweit gelingt es, den untersuchten Gegenstandsbereich in den Daten abzubilden? Sind die auf der Basis des vorliegenden Datenmaterials gefundenen *kausalen Zusammenhänge* oder *Mechanismen* tatsächlich maßgeblich für reale Vorgänge im untersuchten Gegenstandsbereich oder sind sie dem Wirken bislang unerkannter Einflüsse (intervenierende oder Moderatorvariablen, unbeobachtete Heterogenität) zuzuschreiben?

Ein hypothetiko-deduktives Vorgehen bietet im wesentlichen zwei Angriffspunkte für Probleme der internen Validität, d.h. für Probleme, die mit der Theoriebildung bzw. dem Verhältnis zwischen Daten und Theorie zusammenhängen. Den einen Angriffspunkt bilden *Operationalisierung und Indikatorbildung* - das gemessene Phänomen kann eine unangemessene Repräsentation jenes theoretischen Gegenstandes darstellen, über den eine Aussage gemacht werden soll. Der andere Angriffspunkt liegt in der prinzipiellen Vieldeutigkeit empirischer Phänomene: jeder beobachtete Sachverhalt läßt sich nicht nur mit Hilfe einer vorab formulierten Hypothese erklären, sondern auf vielfältige Weise interpretieren. Forschungspraktisch bedeutet dies, daß Zweifel an den im Datenmaterial entdeckten Zusammenhängen *immer möglich* sind, denn stets kann eine bislang unbekannte intervenierende Variable einen theoretisch bedeutsamen Effekt sowohl maskieren als auch künstlich erzeugen. Dieses Problem läßt sich im wesentlichen nur dadurch bearbeiten, daß jeweils konkurrierende Erklärungshypothesen, die gleichermaßen theoretisch plausibel sind, formuliert werden und nach zusätzlicher empirischer Evidenz bzw. Gegenevidenz für die verschiedenen Alternativen gesucht wird. Eine solche Prüfung der internen Validität durch die Formulierung alternativer Erklärungen und die anschließende Suche nach bestätigenden oder widerlegenden Instanzen ist natürlich prinzipiell unabschließbar, da es grundsätzlich nie möglich sein wird, die Existenz von bislang unbekannten und verborgenen Einflüssen, die einen untersuchten Effekt mit verursachen, auszuschließen.

Bei der Entwicklung von Strategien zur Sicherung interner Validität für die qualitative Sozialforschung muß allerdings berücksichtigt werden, daß sich - hinsichtlich des Stellenwertes und der Rolle der Theoriebildung im Forschungsprozeß - das qualitative und interpretative Methodenparadigma grundlegend von hypothetiko-deduktiven Forschungsstrategien unterscheidet. Während theoretische Aussagen dort als Hypothesen vor dem Kontakt mit dem empirischen Feld formuliert werden müssen und die Erhebung und Auswertung der Daten zu deren Überprüfung vorgenommen wird, erfolgt die Theoriebildung im Rahmen einer qualitativen Forschungslogik *ex post* auf der Basis des empirischen Materials. Zur Konzeptualisierung dieses Prozesses wurden oftmals *induktivistische Modelle des Forschungsprozesses* vorgeschlagen: demnach *emergieren* die theoretischen Konzepte quasi von selbst aus dem Datenmaterial, wenn der Untersucher es auf sich nimmt, sich dem Datenmaterial möglichst vorurteilslos zu nähern, d.h. im Einzelfall sogar darauf zu verzichten, die wissenschaftliche Literatur über das Feld zu studieren (vgl. GLASER, STRAUSS 1967, S.37). Die erkenntnistheoretische Unzulänglichkeit eines solchen *tabula rasa* Modells empirischer Forschung, welches auf einer naiven Fehlinterpretation menschlichen Erkenntnisvermögens beruht, hat die moderne Wissenschaftsphilosophie zur Genüge aufgezeigt (vgl. Kap. 1, S.9). Auch in der Forschungspraxis werden die Beschränkungen solcher induktivistischer Konzeptionen deutlich:

Daß induktivistische Modelle letzten Endes überhaupt keine methodischen Verfahrensregeln für qualitative Forschung enthalten, wird deutlich, wenn "Novizen" sich in ihrer Untersuchungstätigkeit auf die von Glaser und Strauss in ihrer 1967 erschienenen Monographie vorgeschlagenen Regeln stützen wollen: bei dem Versuch, möglichst ohne Vorannahmen in das empirische Feld zu gehen und auf das "Emergieren" theoretischer Konzepte zu warten, "ertrinken" sie geradezu im Datenmaterial. Allerdings zeigen die (zur selben Zeit wie die erwähnte Monographie) erschienenen Berichte der beiden Autoren über ihre eigene Forschungspraxis, daß der Induktivismus eine unzutreffende Selbstbeschreibung der eigenen Untersuchungstätigkeit darstellte:

"Zunächst möchten wir erklären, daß unser Konzept des 'Bewußtseinskontextes' durch persönliche Erfahrungen beider Autoren vorgezeichnet war. (...) Kurz nachdem sich Strauss und Glaser zusammengetan hatten, arbeiteten sie systematisch die Konzepte (und Typen) von Todeserwartungen und Bewußtseins-Kontexten sowie das Paradigma für die Untersuchung der Bewußtseinskontexte aus. *So wurde die Erhebung der präliminaren Daten bereits von den Vorstellungen der*

Todeserwartungen und Bewußtheit beeinflusst¹⁰." (GLASER, STRAUSS 1974, S. 264)

Während die Autoren induktivistische *tabula rasa* Konzepte vor allem zur Abgrenzung nach außen gegenüber hypothetiko-deduktiven und "verifikationistischen" Modellen des Forschungsprozesses entwickelten, wurde ihre tatsächliche Praxis von einer ganz anderen Strategie bestimmt: auf der Grundlage (teilweise expliziter, teilweise impliziter) theoretischer Vorannahmen wurde empirisches Material interpretiert und gedeutet. Die theoretischen Vorannahmen nahmen dabei allerdings nicht die Form präzise operationalisierter Hypothesen an, sondern stellten nur einen *heuristisch-analytischen Rahmen* dar, ein Raster, daß die Strukturierung beobachtbarer Phänomene und deren Einordnung in bereits vorhandene Wissensbestände ermöglichte. Im Gegensatz zu den Hypothesen, welche im Rahmen eines deduktiven Forschungsdesigns getestet werden, enthält der heuristische Rahmen in hohem Maße solche theoretischen Aussagen, die einen relativ hohen Allgemeinheitsgrad und einen niedrigen empirischen Gehalt und damit einen geringen Falsifizierbarkeitsgrad aufweisen - dies können bspw. allgemeine handlungs-theoretische Annahmen sein, wie sie in Teilprojekt A1 des Sfb zur Analyse biographischer Konstruktionen verwendet werden (KOCK, WITZEL 1993). *Empirisch gehaltvolle Aussagen* werden erst auf der Basis empirischen Materials formuliert.

Ein solches Vorgehen der empirisch begründeten Theoriebildung ist durchaus vereinbar mit solchen in jüngeren wissenschaftstheoretischen Diskussionen über die Rationalisierbarkeit des *Entdeckungszusammenhangs* und über die Bedeutung von *Methodologien der Entdeckung* im Forschungsprozeß entwickelten Konzepten (vgl. etwa DANNEBERG 1989; FISCHER 1983; NICKLES 1980, 1985, 1990). Diesen Konzepten zufolge ist die Entdeckung von Hypothesen nicht allein durch geniale Einfälle und mystische Intuitionen von Forschern erklärbar, sondern Folge *rationaler Schlußvorgänge*. Die logische Grundlage hierfür bilden allerdings nicht induktive Schlußfolgerungen, sondern ein logischer Schlußmodus, der als *hypothetisches Schließen* (vgl. dazu HANSON 1965)¹¹ bezeichnet werden kann. Hierbei werden

¹⁰ Hervorhebung durch die Autoren.

¹¹ Dieser Begriff geht ursprünglich auf Charles S. Peirce zurück, der verschiedene Konzepte des hypothetischen Schlusses entwickelt hat. Manche Autoren bevorzugen auch die ebenfalls von Peirce vorgeschlagenen Begriffe "*Retroduktion*" oder "*Abduktion*", übersehen dabei aber, daß eine Abduktion (zumindest in Peirce's Spätwerk) nur eine besondere Form des hypothetischen Schlusses darstellt (ausführlich hierzu REICHERTZ 1991).

ausgehend von theoretischem Vorwissen Erklärungen für beobachtete Phänomene *ex post* formuliert, indem diese Phänomene entweder unter vorhandene Kategorien subsumiert werden oder indem neue Kategorien unter Zuhilfenahme vorhandener Wissensbestände (ggf. durch deren Modifikation) konstruiert werden. Dieses Konzept des hypothetischen Schließens ist im Gegensatz zu induktivistischen Modellen *fallibilistisch*: es wird nicht (wie bei induktivistischen Konzepten) davon ausgegangen, daß die Validität der aus dem Datenmaterial gewonnenen Aussagen bereits dann gesichert ist, wenn der Untersucher die Dinge im Untersuchungsfeld "unvoreingenommen" und "vorurteilslos" betrachtet. Vielmehr führen hypothetische Schlußfolgerungen zwar zu rational begründeten (weil mit empirischen Phänomenen und theoretischem Vorwissen kompatiblen) Einsichten. Diese Einsichten sind jedoch in hohem Maße fallibel, was sich in der Regel bereits darin zeigt, daß dasselbe empirische Phänomen oftmals mehrere theoretische Erklärungen zuläßt, die zwar alle mit vorhandenen Wissensbeständen kompatibel sind, sich aber dennoch logisch ausschließen.

Dieser Umstand wird bei der Auswertung qualitativen Datenmaterials dann deutlich, wenn eine Untersucherguppe dasselbe Material bearbeitet. Qualitatives Datenmaterial, in der Regel bestehend aus Interview- oder Beobachtungsprotokollen, läßt oftmals eine ganze Reihe verschiedener Interpretationsmöglichkeiten zu. Zu einer Textstelle können ganz verschiedene und logisch inkompatible Deutungshypothesen konstruiert werden, die gleichermaßen vom empirischen Material gestützt werden. Während dieser Umstand bei einem einzelnen Auswerter verdeckt bleiben kann, können sich innerhalb einer Forschungsgruppe divergierende Lesarten entwickeln, d.h. es werden verschiedene Erklärungshypothesen für dasselbe empirische Phänomen formuliert. Wird das Datenmaterial mit Hilfe eines Kategorienschemas kodiert, äußert sich dieses Problem als dessen mangelnde *Reliabilität*: verschiedene Auswerter ordnen derselben Textstelle unterschiedliche Kodekategorien zu.

Die mit der Analyse qualitativen Datenmaterials verbundenen *threats for validity* sind also den von Cook und Campbell beschriebenen Bedrohungen der internen Validität und der Konstruktvalidität¹² analog: *Jedes empirische Phänomen läßt eine ganze Reihe*

¹² Bei einem qualitativen Forschungsdesign läßt sich somit nur schwer zwischen interner Validität und Konstruktvalidität unterscheiden. Wir betrachten deshalb das Problem, daß diesselben empirischen Phänomene von unterschiedlichen Untersuchern als Ausdruck verschiedener theoretischer Zusammenhänge gedeutet werden können, nicht (wie Cook und Campbell) als eigenes Problem der Konstruktvalidität, sondern betrachten es als Teil der internen Validität.

unterschiedlicher Erklärungen zu, ein Phänomen, welches durch den Einfluß eines Faktors A bislang erklärt wurde, kann ebenso durch einen bislang unbekannten Faktor B bedingt sein.

Als weiterer *threat for validity* kommt die Gefahr von *selektiven* und *ad-hoc Interpretationen* hinzu. Interpretationsmuster können sich verselbständigen, so daß die Interpreten das folgende Datenmaterial nur noch im Licht von einmal formulierten theoretischen Aussagen sehen, während dort auftauchende Gegenevidenz durch *ad-hoc Annahmen* und *immunisierende Hilfhypothesen* weginterpretiert wird.

Eine Reihe von Strategien zur Bewältigung dieser Bedrohungen der internen Validität sind verfügbar, manche in der Literatur beschrieben, manche stellen einen Teil jener von Katz und Festinger beschriebenen *Folklore der Sozialwissenschaften* (FRIEDRICHS 1980, S.376) dar - dies sind im Kontext qualitativer Sozialforschung häufig verwendete Strategien, deren methodologische Bedeutung jedoch nur selten reflektiert werden. Im folgenden werden verschiedene Strategien zur Sicherung der internen Validität diskutiert: die diskursive Validierung bzw. die nominale Gruppentechnik, verschiedene Methoden zur Überprüfung von Deutungshypothesen und EDV-gestützte Verfahren der Archivierung und Analyse von qualitativen Daten.

4.1.2 Diskursive Validierung und nominale Gruppentechnik

Verfahren zur Bewältigung der angesprochenen Bedrohungen der internen Validität erfordern ein Vorgehen, bei dem eine möglichst große Anzahl von Erklärungshypothesen expliziert und gegeneinandergehalten werden kann. Die Ausschöpfung aller möglichen Erklärungen, Lesarten und Deutungshypothesen für eine Textstelle ist ein für die hermeneutische Sozialforschung häufig empfohlenes Vorgehen. So schlagen etwa Heinze und Klusemann diese Strategie vor, weil bei der Interpretation qualitativer Interviews die für die Alltagskommunikation üblichen "Validierungsverfahren" fortfallen: der Umstand, daß die Interaktionspartner einander verstehen, ist nicht aus deren Handeln ablesbar:

"Wir befinden uns mit der Interpretation in einer von Alltagskommunikation abgehobenen Situation insofern, weil in dem Verhältnis zwischen uns als Forscher und den Interviewten ein Handlungszwang entfällt. Handlung ist aber eine wesentliche Rückmeldung für Interaktion. Da wir aber von den Interviewten zunächst keine Rückmeldung haben, müssen wir alle möglichen Bedeutungskomponenten aus den Protokollen erschließen. Das heißt, wir versuchen, alle Bedeutungsgehalte, die eine Äußerung haben kann, aufzulisten und betrachten diese Gehalte zunächst prinzipiell als gleichwahrscheinlich und gleichwertig." (HEINZE, KLUSEMANN 1980, S.101).

Ein solches Vorgehen der *Exhaustation von Deutungshypothesen* zählt auch zu den grundlegenden Auswertungsschritten der "objektiven Hermeneutik" (OEVER-MANN u.a. 1979): Unterschiedliche Interpretationen einer Textstelle werden in Diskussionen innerhalb der Forschergruppe gegeneinandergehalten. Dabei geht es vor allem um divergierende Erklärungshypothesen, die auf der Grundlage eines argumentativen Prozesses beibehalten, aufgegeben oder angeglichen werden sollen. Die Intersubjektivität der Interpretationsergebnisse soll dadurch garantiert werden, daß eine Gruppe von Forschern regelgeleitet und nachvollziehbar über Interpretationen diskutiert. Ziel ist dabei, möglichst *alle* denkbaren und theoretisch sinnvollen Erklärungshypothesen zu äußern und kritisch zu würdigen.

Nur wenige qualitative Forscher führen allerdings extensive Feinanalysen ihres Datenmaterials nach der Methode der "objektiven Hermeneutik" durch. Ein Vorgehen, bei dem verschiedene Auswerter unabhängig voneinander dasselbe Dokument bearbeiten und anschließend die verschiedenen Ergebnisse gemeinsam diskutieren, ist

als Validierungsverfahren in der qualitativen Sozialforschung jedoch weit verbreitet. Ein solches, als "*nominale Gruppentechnik*" bezeichnetes Verfahren läßt sich mit den von Oevermann und Kollegen oder von Heinze und Klusemann vorgeschlagenen Strategien zur Exhaustation von Deutungshypothesen zur Gruppe der *diskursiven Validierungsmethoden* zusammenfassen, deren Ziel es ist, die *Gültigkeit* von Interpretationen durch Diskussionen in der Forschungsgruppe zu überprüfen.

Ein Konsens über Deutungshypothesen kann jedoch keineswegs als alleiniges Gütekriterium für qualitative Deutungshypothesen dienen (wie dies etwa bei "konsens-theoretischen Wahrheitskonzepten" der Fall wäre, die Kvale (1989, 1991) als Grundlage von Gütekriterien für die qualitative Sozialforschung empfiehlt). Hierdurch würde die Zustimmung zu einer bestimmten Hypothese vollständig abhängig vom gemeinsam geteilten theoretischen Vorwissen der Interpreten. Dies kann zwei äußerst problematische Konsequenzen nach sich ziehen, je nachdem ob die Forschungsgruppe hinsichtlich bestimmter theoretischer Vorannahmen, Vorurteile oder Vorlieben sehr heterogen oder sehr homogen ist:

- * Eine theoretisch sehr homogene Forschungsgruppe kann sehr schnell Einigung über die Bedeutung eines bestimmten Vorfalles oder einer bestimmten Textstelle erzielen, ohne daß hierdurch eine Aussage über deren Gültigkeit getroffen werden kann. Ein ironisches Gedankenexperiment des englischen Sozialhistorikers Trevor-Ropper mag dies verdeutlichen:

"I can easily imagine two grave old demonologists chatting comfortably by a porcelain stove in 17th century Würtemberg or Bavaria and discussing, over a flagon of Bocks-beutel, the case of a suspicious old lady in a remote village. On a few scraps of gossip, insignificant to the layman but full of meaning to the expert - the furtive comings and goings of a black cat, the wasting away of one's parson's piglings, an alleged rustle in the chimney stack - they both reach the same conclusion." (Trevor-Ropper 1973)

- * Ist die Forschungsgruppe theoretisch sehr heterogen, so kann eine Debatte über die Bedeutung einer einzelnen Textstelle *ad nauseam* ausgedehnt werden, ohne daß eine Einigung über eine angemessene Interpretation erzielt wird.

Die Selbstblockade einer theoretisch heterogenen Forschungsgruppe kann nur dadurch verhindert werden, daß bei der diskursiven Validierung bestimmten Regeln gefolgt wird: es müssen Kriterien festgelegt werden, die angeben, wodurch eine Auseinandersetzung über divergierende Interpretationen entschieden werden kann. Um die Gefahr abzuwenden, daß eine Forschungsgruppe sich - ähnlich wie die Dämonologen in dem von Trevor-Ropper angestellten Gedankenexperiment - aufgrund ihrer theoretischen Homogenität vorschnell über die Gültigkeit spekulativer Annahmen verständigt, müssen solche Regeln allerdings zweierlei sicherstellen:

- * Die formulierten Deutungshypothesen müssen *empirisch gehaltvoll* bzw. *falsifizierbar* sein, eine Eigenschaft, die auf die Deutungshypothesen von Dämonologen grundsätzlich nicht zutrifft - hier ist i.d.R. kein empirisches Ereignis denkbar, welches die Hypothese widerlegen könnte, daß es sich bei der inkriminierten Frau um eine Person handelt, die sich der schwarzen Magie bedient.
- * Es muß möglich sein, widersprechende Evidenz tatsächlich im Datenmaterial aufzufinden.

Die Formulierung potentieller Falsifikatoren und die Suche nach empirischer Gegenevidenz ließen sich als die zentralen Abgrenzungskriterien betrachten zwischen *methodisch kontrolliertem Fremdverstehen* und solchen hermeneutischen Bemühungen, wie sie von den erwähnten Dämonologen betrieben werden.

Die interpretative Sozialforschung muß sich hier mit einer besonderen Bedrohung der internen Validität auseinandersetzen: Bei einem qualitativen Design, welches auf einer Methodologie empirisch begründeter Theoriekonstruktion beruht, können Deutungshypothesen nicht experimentell überprüft werden, sondern empirische Gegenevidenz muß in bereits vorhandenem Datenmaterial *ex post* gesucht werden. Dadurch wird die *Falsifikation von Hypothesen unwahrscheinlicher, je spezifischer sie formuliert sind*. Im Folgenden sollen Strategien zur Überprüfung von Deutungshypothesen im qualitativen Forschungsprozeß diskutiert werden, mit denen diesem *threat for validity*, den man als *Validierungsparadox interpretativer Datenanalyse* bezeichnen kann, begegnet werden kann.

4.1.3 Die empirische Überprüfung von Deutungshypothesen am Material

Von entscheidender Bedeutung für eine Überprüfung von Deutungshypothesen ist, daß systematisch nach Evidenz *und* Gegenevidenz im Datenmaterial gesucht wird. Das angewendete Verfahren zur internen Validierung muß also verhindern, daß einmal formulierte Hypothesen im weiteren Verlauf der Auswertung immunisiert werden, indem etwa nur die bestätigenden Instanzen im Datenmaterial registriert werden. Das bedeutet, daß bei der Überprüfung von Deutungshypothesen unbedingt *alle* für die jeweilige Hypothese relevanten Textstellen berücksichtigt werden. Dies läßt sich natürlich am einfachsten sicherstellen durch eine *sequentielle Analyse des gesamten Datenmaterials*.

4.1.3.1 Sequentielle Analyse

Solch eine sequentielle Analyse ist die Grundlage von zwei (zumindest in Deutschland) weit verbreiteten Verfahren interpretativer Sozialforschung: der "*strukturel-len Auswertung*" *narrativer Interviews* nach Schütze und der *objektiven bzw. strukturalen Hermeneutik* nach Oevermann (vgl. BOHNSACK 1991, S.35). In der objektiven Hermeneutik wird etwa betont, daß nach der "*extensiven Explikation der von den anfänglichen Interakten gedeckten möglichen Interpretationen*" (OEVERMANN ET AL. 1979, S.401) bei der weiteren sequentiellen Bearbeitung des Materials versucht wird "*sukzessive Inkonsistenzen und falsche Interpretationsansätze auszuschließen*" (ebd.). In einem solchen Vorgehen sehen Oevermann und Kollegen "*das hermeneutische Analogon zur Popper'schen Idee der Falsifikation von Hypothesen*" (ebd., S.391). Auf das bereits erwähnte *Validierungsparadox* bei der interpretativen Auswertung von verbalen Daten gehen diese Autoren allerdings nicht ein. Dabei stellt sich dieses Problem gerade im Rahmen der objektiven Hermeneutik, die als Interpretationsverfahren nur zur Analyse einiger weniger Sätze eines Dokumentes, einer Interaktion von wenigen Minuten Dauer, oder den ersten Seiten eines Interviewtranskriptes geeignet ist: Die Wahrscheinlichkeit, falsifizierende Instanzen zu finden, ist bei wenig umfangreichem empirischen Material natürlich entsprechend gering.

Auch das sequentielle Vorgehen bei der strukturellen Beschreibung narrativer Interviews läßt i.d.R. nur die Analyse einiger weniger Fälle mit vertretbarem Zeitaufwand zu. Ein striktes sequenzanalytisches Vorgehen ist also nur dann möglich, wenn der

Umfang des Datenmaterials entsprechend eingeschränkt wird. Dies wiederum kann erhebliche Probleme der *externen Validität* mit sich bringen. Bei qualitativen Auswertungsverfahren, bei denen umfangreichere Datenmengen - etwa die Protokolle mehrmonatiger Feldbeobachtungen oder mehr als zehn offene Interviews - erhoben werden, kann deshalb eine sequentielle Feinanalyse nur zu Beginn des Auswertungsprozesses durchgeführt werden, wie dies bspw. Glaser für die erste Auswertungsphase bei der Anwendung der Methodologie empirisch begründeter Theoriebildung ("*grounded theory*") empfiehlt (GLASER 1978, S.57). Eine systematische sequentielle Suche nach empirischer Evidenz oder Gegenevidenz, die auch in späteren Auswertungsphasen i.d.R. mehrfach durchgeführt werden muß, bringt bei umfangreichen Datenmengen dann allerdings nicht zu bewältigende forschungsökonomische Probleme mit sich. Schließlich ist der *data overload* ein in der qualitativen Sozialforschung ohnehin häufig beklagtes Problem (MILES, HUBERMAN 1993, S.55 f.). Um der Gefahr zu begegnen, daß dann, wenn sequentielle Verfahren zu aufwendig sind, einzelne, willkürlich herausgesuchte Textstellen zur "Stützung" (de facto jedoch nur zur empirischen Illustration) von Hypothesen verwendet werden, muß das Datenmaterial in einer spezifischen Weise strukturiert werden, die es ermöglicht, systematisch Gegenevidenz im Datenmaterial zu entdecken.

4.1.3.2 Qualitative Kodierung und Hypothesenprüfung

Die grundlegende Operation zur Strukturierung großer Mengen qualitativen Datenmaterials stellt die *qualitative Kodierung* dar (vgl. GLASER 1965, 1978; STRAUSS, CORBIN 1990; MILES, HUBERMAN 1993, S.54 ff., TESCH 1990): bestimmten Textstellen werden Qualifikatoren (Kodes oder theoretische Kommentare) zugeordnet¹³, wobei die Kodes entweder die Kategorien eines zuvor erstellten (evtl. modifizierbaren) Kategorienschemas darstellen (MILES, HUBERMAN 1993, S.57) oder - beim "*offenen Kodieren*" (GLASER 1978, S. 54; STRAUSS, CORBIN 1990, S.61 ff.) - *ad hoc* auf der Basis der im Material vorhandenen Begrifflichkeiten der Beforschten gebildet werden. Diese Kodierung läßt sich mit einem *permanenten Vergleich* der kodierten Textpassagen verbinden: "*While coding an incident for a category, compare it with the previous incidents in the same and different groups coded in the same category.*" (GLASER, STRAUSS 1967, S. 106) Durch den permanenten Vergleich können - so Glaser und Strauss - Kategorien und ihre Subkategorien und Dimensionen aus dem Datenmaterial entwickelt werden. Die vergleichende Analyse von kodierten Textpassagen läßt sich aber nicht nur als *heuristisches Verfahren zur Hypothesengenerierung*, sondern auch als Methode zur *Überprüfung qualitativer Deutungshypothesen* einsetzen. Hierbei können zwei Wege eingeschlagen werden:

- * Bei einer globalen Suche nach Evidenz bzw. Gegenevidenz im Datenmaterial müssen alle für eine bestimmte Hypothese relevanten Textstellen (i.d.R. sind dies alle Textsegmente, welche mit einer bestimmten Kategorie kodiert worden sind) gesucht und analysiert werden.

So wurde bspw. bei einer am Sfb durchgeführten Auswertung von Interviews mit betrieblichen Ausbildern die Hypothese formuliert, daß bei der Bewerberauswahl für betriebliche Ausbildungsplätze die Persönlichkeitseigenschaften und soziale Kompetenzen der Bewerber eine wesentlich größere Rolle spielen als deren Schulnoten. Zur Überprüfung dieser Hypothesen wurden alle Textpassagen in den Interviews ausfindig gemacht, welche mit dem Kode "Bewerberauswahl" versehen waren.

¹³ Die Kodes bzw. theoretischen Kommentare werden üblicherweise am rechten oder linken Rand des Textes oder auf Karteikarten vermerkt (GLASER, STRAUSS 1967, S.106).

- * Bei einer Suche nach *entscheidenden Fällen* oder *entscheidenden Textpassagen*, die als potentielle Falsifikatoren für eine Deutungshypothese dienen können, muß die Suche nach Textpassagen durch bestimmte Kriterien eingeschränkt werden.

Bei der weiteren Analyse der erwähnten Ausbilderinterviews wurde die Hypothese tentativ formuliert, daß das Auswahl-kriterium "sozialisatorisch kompetente Herkunftsfamilie" nur für die Bewerberauswahl in Kleinbetrieben von Bedeutung ist. Zur Überprüfung dieser Hypothese wurde nach *kontra-stierenden Textpassagen* - konkret: nach Textsegmenten mit dem Kode "Bewerberauswahl" in Interviews mit Ausbildern von Großbetrieben - gesucht.

Methoden der Fallkontrastierung werden in der gegenwärtigen Debatte insbesondere als Heuristik zur Hypothesengenerierung (vgl. GLASER, STRAUSS 1967, S.45 ff.) oder der Typenbildung beschrieben (GERHARDT 1986, S.69f.; 1991, S.438). Für eine ausführliche Darstellung ihrer Funktion als Verfahren zur Überprüfung qualitativer Hypothesen muß auf ältere Arbeiten aus dem Kontext der *Chicagoer Schule* der naturalistischen Feldforschung zurückgegriffen werden. Unter dem Begriff "*analytische Induktion*" wird hier ein Vorgehen beschrieben, bei dem sich der Untersucher oder die Untersucherin mit einer tentativen Hypothese dem Untersuchungsfeld nähert und dort systematisch nach *entscheidenden Fällen* ("*crucial cases*"), die als potentielle Falsifikatoren dienen können, sucht (LINDESMITH 1968; CRESSEY 1950, 1971; kritisch ROBINSON 1951, 1952). Die entscheidenden Fälle sollen dabei als *Analogon* der *experimenta crucis* in den experimentellen Wissenschaften eine empirisch begründete Entscheidung über die Geltung der anfänglich formulierten Hypothesen zulassen. Empirische Gegenevidenz führt dazu, daß entweder die Ausgangshypothese modifiziert, das erklärende Phänomene undefiniert oder das zu untersuchende Phänomen ausgeschlossen werden muß. Dabei muß jede Modifikation oder Umdefinition zu einer Präzision der Ausgangshypothese und des Untersuchungsgegenstandes führen.

Die Suche nach entscheidenden Fällen und die Modifikation von Ausgangshypothesen wird solange fortgesetzt, bis keine der Hypothese widersprechenden Fälle mehr entdeckt werden können. Auf diese Weise kann nach der Untersuchung einer kleinen Zahl von Fällen praktische Sicherheit erreicht werden. Die Entdeckung eines einzigen negativen Falles durch den Untersucher oder einen anderen Forscher widerlegt jedoch bisherige

theoretische Erklärungen und verlangt deren Reformulierung. "(...) *this procedure of examining cases, re-defining the phenomenon and re-formulating the hypothesis is continued until a universal relationship is established, each negative case calling for re-definition or re-formulation.*" (CRESSEY 1971, S. 16)

Bei einem solchen Vorgehen, bei dem die entstehende Theorie sukzessive in Auseinandersetzung mit dem Datenmaterial entwickelt wird, wird eine weitere bereits erwähnte Bedrohung der internen Validität virulent: Theoretische Konstruktionen und darauf beruhende Interpretationsmuster können sich so weitgehend verselbständigen, daß im Datenmaterial auftauchende Gegenevidenz durch *ad-hoc Annahmen* und *immunisierende Hilfhypothesen* uminterpretiert wird.

Methodologische Regeln, um diesem *threat for validity* zu begegnen, lassen sich unter Rückgriff auf jene *Modelle rationaler Theorietransition* formulieren, die in der modernen Wissenschaftstheorie in kritischer Bezugnahme auf Karl Poppers Konzept des Erkenntnisfortschritts von Imre Lakatos und Larry Laudan formuliert worden sind (LAKATOS 1982; LAUDAN 1977). Diese Modelle beruhen auf einer Unterscheidung zwischen *progressiven* und *degenerativen Problemverschiebungen*:

Voraussetzung für eine progressive Problemverschiebung ist Lakatos (1982) zufolge die Zunahme an empirischem Gehalt: Eine neue Theorie muß zur Aufklärung jener Anomalien führen, die zum Scheitern der Vorgängertheorie geführt haben und über die Phänomene hinaus, die diese erklärt, weitere empirisch beobachtbare Sachverhalte erklären und prognostizieren können. Die Immunisierung bestehender Theorien durch die Einführung gehaltsvermindernder *ad-hoc-Hypothesen* führt demgegenüber zu degenerativen Problemverschiebungen. Laudan (1977) führt ein weiteres Kriterium für eine progressive Problemverschiebung ein. Die Modifikation einer Theorie stellt auch dann einen Fortschritt dar, wenn die neue Theorie weniger konzeptuelle Probleme besitzt als ihre Vorgängerin, wobei konzeptuelle Probleme einerseits in internen Inkonsistenzen und andererseits in der mangelnden Kompatibilität mit anderen bislang bewährten Erklärungsansätzen besteht. Hierdurch verdeutlicht Laudan, daß - neben der Einführung *gehaltvermindernder Hilfhypothesen* - auch zwei weitere Strategien bei der Benutzung von *ad-hoc-Hypothesen* zu einer degenerativen Problemverschiebung führen können: (1.) die Verwendung von Hilfhypothesen, die an eine falsifizierte Theorie "angeklebt" werden, zu der sie in keinem inhaltlichen Zusammenhang stehen und (2.)

die Verwendung von Hypothesen, die zu anderen, bislang gut bewährten Theorien einen Widerspruch erzeugen.

Um eine progressive Problemverschiebung einzuleiten, muß eine neue Hypothese also mindestens eines der drei folgenden Kriterien erfüllen:

1. Sie muß zu einem Zuwachs an empirischem Gehalt führen, d.h. die durch die neue Hypothese veränderte Theorie muß zusätzliche Tatsachen aufklären oder neue empirische Sachverhalte voraussagen.
2. Sie muß zu einer Verminderung interner konzeptueller Probleme durch einen Zuwachs an logischer Kohärenz, innerer Geschlossenheit oder Sparsamkeit führen.
3. Sie muß die Anschlußfähigkeit der Theorie, in die sie eingebunden ist, erhöhen, d.h. die zentralen Aussagen der (durch die Hypothese veränderten) Theorie müssen sich mit anderen bislang bewährten Theorien für diesen Gegenstandsbereich verknüpfen lassen.

Hierbei muß allerdings beachtet werden, daß nicht eine Eigenschaft zu Lasten einer anderen verbessert werden darf: ein wachsender empirischer Gehalt einer Theorie darf also nicht durch eine verminderte Konsistenz "erkauft" werden.

Die folgenden Überlegungen zu forschungstechnischen und forschungsökonomischen Aspekten der skizzierten Validierungsstrategien sollen die Behandlung dieses Themas abschließen:

Die manuellen Verfahren, die für eine Kodierung und Strukturierung qualitativen Datenmaterials verwendet werden, sind, trotz ihrer erheblichen methodologischen Bedeutung, nur selten in der Literatur beschrieben worden (Ausnahme: MILES, HUBERMAN 1993, S.54 ff.; TESCH 1990, S.113 ff.): sie umfassen *intellektuelle Handwerkstechniken* wie den Aufbau von umfangreichen Karteien, Zettelkästen und Fundstellenregistern. Auch wenn das qualitative Datenmaterial in einer solchen Weise aufbereitet und strukturiert wurde, kann die Suche nach empirischer Evidenz oder Gegenevidenz einen zeitraubenden Prozeß darstellen. Eine entscheidende Innovation stellt deshalb die Einführung EDV-gestützter Systeme zur Verwaltung und Archivierung qualitativen Datenmaterials dar (LEE, FIELDING 1991; TESCH 1990; KELLE 1990, 1992), weil die Suche nach relevanten Textsegmenten hierdurch wesentlich abgekürzt werden

kann. Im Gegensatz zu vielen manuellen Techniken steht beim EDV-gestützten Vorgehen zudem immer der Originalkontext der jeweiligen Textsegmente zur Verfügung. Nur aus Gründen der Forschungsökonomie muß bei weiteren Auswertungsschritten wie etwa der Entwicklung von Typologien somit nicht auf Zusammenfassungen zurückgegriffen werden, da Zusammenhänge mit geringem Aufwand am Originaltext expliziert werden können. In jüngerer Zeit sind eine Reihe von EDV-gestützten Systemen der Textverwaltung speziell im Rahmen qualitativer Forschungsprojekte entwickelt worden - diese stellen jedoch nur Werkzeuge zur Strukturierung und Verwaltung des Datenmaterials dar, die *interpretative Auswertung* bleibt Aufgabe des Forschers, dessen Zugriff auf das Datenmaterial jedoch wesentlich erleichtert wird. Hierbei wird das Vorgehen bei der Analyse - insbesondere umfangreicherer - qualitativer Datenbestände systematisiert und transparenter gestaltet, wobei die einzelnen Analyseschritte der Textanalyse, Typen- und Theoriebildung offengelegt werden und damit sowohl innerhalb der Forschungsgruppe als auch für außenstehende Forscher nachvollziehbarer und überprüfbarer werden.

4.2 Verfahren zur Sicherung der externen Validität

4.2.1 Bedrohungen der externen Validität in der qualitativen Sozialforschung

Guttman und Zetterberg bezogen den Begriff der externen Validität auf die Möglichkeit, Untersuchungsergebnisse zur Grundlage von *Prognosen* zu machen. Eine hohe statistische Korrelation zwischen dem Ergebnis von Tests und Befragungen einerseits und einem Außenkriterium andererseits wären demnach als Zeichen für eine hohe externe Validität zu werten. Bei beiden Autoren wurde damit vor allem der technische Aspekt dieses Konzepts im Kontext von Einstellungsmessung und Surveymethodologie angesprochen. Allerdings läßt sich der Hinweis von Zetterberg auf die Gefahr von Verfälschungen und Verzerrungen, die durch das Bedürfnis des Interviewten nach sozialer Anerkennung entstehen, bereits als Hinweis darauf betrachten, wie dieses Konzept auf andere Forschungsmethodologien verallgemeinerbar wäre.

Cook und Campbell haben sich schließlich um eine allgemeine Definition dieses Begriffs bemüht: "*We shall use the term external validity to refer to the approximate validity with which conclusions are drawn about the generalizability of a causal relationship to and across populations of persons, settings, and time.*" (COOK, CAMPBELL 1979, S.39) Forschungsergebnisse, die sich über den Rahmen der Untersuchungssituation und Untersuchungspopulation auf zahlreiche soziale Situationen und Populationen hinaus generalisieren lassen, besitzen hohe externe Validität. Die Frage, inwieweit sich Untersuchungsergebnisse über die konkrete Untersuchungssituation und die jeweilige Untersuchungsgruppe verallgemeinern lassen, muß grundsätzlich auch in qualitativen Forschungsdesigns gestellt werden. Entsprechend wurden von qualitativen Methodologen, die sich mit Validitätskonzepten für die qualitative Sozialforschung befaßt hatten, nahezu gleichlautende Kriterien definiert, wobei der Begriff der externen Validität entweder direkt übernommen wurde (wie von LeCompte und Preissle (1993)) oder durch Begriffe wie *Übertragbarkeit*, *Vergleichbarkeit* etc. (wie von Lincoln und Guba (1985)), ohne daß hiermit jedoch ein grundsätzlich anderer Aspekt der Validität angesprochen wurde (vgl. Abschnitt 2.2 und 2.3).

Zwei Gefahren können im Kontext einer qualitativen Studie für deren *Verallgemeinerungsfähigkeit* virulent werden:

1. Die Gefahr einer *einseitigen und beschränkten Auswahl von Untersuchungseinheiten*. Fragen der Stichprobenziehung werden in der qualitativen Sozialforschung oftmals, wie Fleck konstatiert, salopp beiseite geschoben (FLECK 1992, S. 758). Die Frage, inwieweit die untersuchten Fälle eine bis zu einem gewissen Grade verallgemeinerungsfähige Auswahl darstellen, kann nicht mit dem Hinweis abgewehrt werden, es handle sich nur um eine "deskriptive" Studie. Daß auch deskriptive Untersuchungen einen zumindest impliziten Anspruch auf Generalisierbarkeit erheben, wird deutlich, wenn man sich vor Augen führt, daß eine deskriptive Studie, die eine völlig verzerrte Stichprobe zum Gegenstand hat, nicht nur völlig wertlos, sondern sogar irreführend sein kann.¹⁴ Auch eine deskriptive Studie bedarf einer sorgfältigen Auswahl der Untersuchungseinheiten bzw. "Stichprobenziehung": *"Sampling and selection are crucial to establishing the authenticity of descriptive analysis because (...) they are the means by which the re-searcher systematically seeks and discards alternative descriptions or explanations of the phenomena observed. They are critical to the logical inferences that support comparability and applicability"*. (LECOMPTE, PREISSLE 1993, S.68)

In bestimmten Traditionen qualitativer Sozialforschung wird der Frage nach Kriterien der Samplebildung auch mit dem Hinweis darauf begegnet, daß sog. "latente Sinnstrukturen", welche soziales Handeln bestimmen, sich in jedem Einzelfall Geltung verschaffen und deshalb allgemeine Strukturgesetze auch anhand von Einzelfällen rekonstruierbar seien (vgl. bspw. OEVER-MANN ET AL. 1979). Eine solche Annahme wäre allerdings nur bei einem hypothetiko-deduktiven (lt. Oevermann "*subsumtionslogischen*") Vorgehen unproblematisch: wenn die in dem Untersuchungsfeld geltenden Strukturgesetze bereits vor der Untersuchung bekannt sind, können diese Annahmen anhand von Einzelfällen verifiziert oder falsifiziert werden. Eine qualitative Forschungsstrategie erfordert demgegenüber, daß die behaupteten Strukturgesetzmäßigkeiten in den Einzelfällen nicht nur *wiedergefunden*, sondern dort systematisch *entdeckt* werden. Dies wäre aber nur dann möglich, wenn man davon ausgehen kann, daß

¹⁴ Als Beispiel dafür können die (eher populärwissenschaftlichen) Untersuchungen des Autorenpaares Benard und Schlaffer dienen: aufgrund von Interviews mit Angehörigen von Berufen mit übermäßiger zeitlicher Belastung (wie bspw. freiberuflichen Rechtsanwälten, Manager etc.) im Alter von 30 bis 40 Jahren ziehen sie weitreichende Konsequenzen über das beschränkte Engagement junger Väter in der Kindererziehung.

der Einzelfall in allen Aspekten vollständig von Strukturgesetzen determiniert wäre. Wird diese ungemein starke Annahme nur in Teilbereichen relativiert, so verfängt sich eine am Einzelfall orientierte qualitative Methodologie sofort in einem unlösbaren "*Zirkel des Verstehens*" (vgl. STEGMÜLLER 1974). Denn nur auf der Basis eines subsumtionslogischen Vorgehens würden sich substantielle Aspekte des Falles (also solche, in denen sich Strukturgesetze äußern) von akzidentiellen oder zufälligen Aspekten unterscheiden, anderenfalls könnte keinerlei Entscheidung darüber getroffen werden, welche Aspekte Ausdruck eines Strukturgesetzes sind und welche nicht.

2. Die zweite schwerwiegende Gefahr für die externe Validität qualitativer Forschungsergebnisse stellen *unkontrollierte Einflüsse* dar, *die von der Untersuchungssituation selber ausgehen*, Einflüsse, die etwa mit dem sozialen Kontext, der durch die Untersuchungssituation erzeugt wird, in Zusammenhang stehen. Der soziale Kontext der Interviewsituation (mit einem Fremden, einem Forscher etc.) kann ein Antwortverhalten provozieren, welches die dem Handeln der Interviewten zugrundeliegenden Deutungsmuster, Motivstrukturen und Optionen nur ungenügend wiedergibt. In der quantitativen Methodenliteratur werden Einflüsse, die von der Untersuchungssituation ausgehen, unter den Stichworten *Artefaktforschung* und *Sozialpsychologie des Experiments* diskutiert: *soziale Erwünschtheit*, *Antworttendenzen* (response sets), *Versuchsleitereffekte* (Rosenthal-Effekt, Hawthorne-effekt usw.) oder *Effekte durch die Anordnung des Untersuchungsmaterials* (Halo-Effekt). Ebenso wie Samplingproblemen wird Problemen, die sich aus der Gestaltung der Untersuchungssituation ergeben, in der qualitativen Sozialforschung oftmals nur wenig Aufmerksamkeit geschenkt, da "*auch qualitative Sozialforscher geneigt sind, dem illusionären 'naturalistischen' Ideal der von Versuchsleitereinflüssen unverzerrten Protokollierung von Beobachtungen anzuhängen.*" (FLECK 1992, S. 754). Die mangelnde Berücksichtigung dieses Problems tritt bspw. in jenen Zweigen der Biographieforschung zutage, die das *narrative Interview* zum Königsweg der Datenerhebung gemacht hat. Hier wird vielfach die Ansicht vertreten, wie Gerhardt moniert, "*daß das Narrative eo ipso als valide bei Interviewdaten zu sehen sei, zufolge phänomenologisch begründeter Setzungen über das Verhältnis von Ereignis und Erinnerung*" (GERHARDT 1985, S. 231). Gerhardt weist darauf hin, daß auch die biographische Stegreiferzählung als Interaktionssituation zu sehen ist, deren Gestaltung bestimmten gesellschaftlich

präformierten normativen Mustern gehorcht. Auch die Erzählsituation im biographischen Interview sei deswegen "*eine Situation ad quem. Sie verfolgt den Zweck, einen Gesprächspartner von etwas zu überzeugen, ihm etwas Glaubhaftes glaubwürdig zu erzählen.*" (ebd., S.242)

Im folgenden sollen einige Validierungsstrategien diskutiert werden, die etwas zur Lösung dieser Methodenprobleme in der qualitativen Sozialforschung und damit zur externen Validität einer qualitativen Studie beitragen können.

4.2.2 Sicherung der Stichprobenvalidität

4.2.2.1 Das Postulat der Verzerrungsfreiheit und das Prinzip der kriterien-geleiteten Stichprobenziehung

Auswahlprobleme beziehen sich auf das Postulat möglichst unverzerrter Stichproben, ein Postulat, welches unzulässigerweise oft einfach mit der Forderung nach *Repräsentativität* gleichgesetzt wird. Dabei wird übersehen, daß Forschungen, die nicht die Sozialberichterstattung, sondern die Konstruktion und Überprüfung von Theorien zum Ziel haben, nur im eingeschränkten Sinn repräsentative Stichproben benötigen (vgl. ZETTERBERG 1965). So muß bei einer Untersuchung, die den systematischen Vergleich zwischen Angehörigen bestimmter sozialer Gruppen anzielt, keineswegs das quantitative Verhältnis dieser Gruppen in der gesamten Bevölkerung berücksichtigt werden. Statt einer Repräsentativerhebung kann die Ziehung der Stichproben bei einem *quasi-experimentellen* Design so erfolgen, daß jede der Untersuchungsgruppen gleich stark vertreten ist. Trotzdem müssen verzerrende Einflüsse vermieden werden, allerdings nur solche Verzerrungen, die theoretisch bedeutsam sind, d.h. die den zu untersuchenden Effekt vortäuschen, verstärken oder maskieren. Ausschlaggebend für die Stichprobenvalidität ist also nicht Repräsentativität schlechthin, d.h. Repräsentativität bezogen auf alle denkbaren Merkmale, sondern *theoriebezogene Repräsentativität*, die sich auf eine bestimmte Anzahl soziologisch bedeutsamer Merkmale beschränkt. Ist eine Stichprobe *theoriebezogen repräsentativ*, so bedeutet dies, daß Träger *theoretisch relevanter Merkmalskombinationen* in der Stichprobe hinreichend vertreten sind, wohingegen solche Merkmale, die für die Fragestellung nicht relevant sind, vernachlässigt werden. Zur Sicherung der externen Validität müssen also

theoretisch signifikante Stichprobenverzerrungen vermieden werden. In der quantitativen Sozialforschung bemüht man sich, das Postulat unverzerrter, d.h. theoretisch repräsentativer Stichproben dadurch zu erfüllen, daß *Zufallsstichproben* gezogen werden. Hierdurch soll sichergestellt werden, daß die Stichprobe hinsichtlich *möglichst vieler, auch unbekannter* Merkmale unverzerrt ist. Bei der Ziehung kleiner Stichproben, wie sie für die qualitative Sozialforschung typisch sind, ist ein solches Vorgehen jedoch problematisch. Hierbei können nämlich die *zufälligen Stichprobenfehler*, die bei großen Samples kaum ins Gewicht fallen, zu folgenschweren Verzerrungen führen. Die oftmals getroffene Feststellung, repräsentative Zufallsauswahlen seien ausreichend für die Konstruktion valider, d.h. unverzerrter Stichproben trifft also nur äußerst eingeschränkt zu. Bei der Ziehung kleiner Stichproben erfordert eine rationale Samplingstrategie, daß durch ein entsprechendes *kriteriengesteuertes Ziehungsverfahren* sichergestellt wird, daß Träger relevanter Merkmalskombinationen berücksichtigt werden. Verschiedene Varianten eines solchen Ziehungsverfahrens, (wie es i.ü. auch bei experimentellen Designs eingesetzt wird (vgl. COOK, CAMPBELL 1979, S.75)) können eingesetzt werden, um die Berücksichtigung theoretisch relevanter Subpopulationen sicherzustellen:

- * Durch eine *maximale Variation der Fälle* (LECOMPTE, PREISSLE 1993, S.72) bzw. *bewußt heterogene Stichprobenauswahl* (COOK, CAMPBELL 1979, S.76) soll erreicht werden, daß das Sample hinsichtlich theoretisch bedeutsamer Merkmale eine möglichst hohe Varianz aufweist. Sonderformen dieses Ziehungsverfahrens stellen die Auswahl von *Extremfällen* oder die Auswahl von *ungewöhnlichen Fällen* (LECOMPTE, PREISSLE 1993, S.75) dar.

- * Bei der Auswahl von Angehörigen von *Modalkategorien* (COOK, CAMPBELL 1979, S.77) bzw. von *typischen Fällen* (LECOMPTE, PREISSLE 1993, S.72) werden Personen mit häufig vorkommenden Merkmalskombinationen in die Stichprobe aufgenommen.

Beide Verfahren können miteinander kombiniert werden, um sicherzustellen, daß eine möglichst große Varianz erzielt wird, ohne daß relativ häufige Merkmalskombinationen vernachlässigt werden. Verzerrungen der statistischen Repräsentativität können dabei bewußt in Kauf genommen werden, indem Angehörige bestimmter Subpopulationen überrepräsentiert werden.

4.2.2.2 Theoretical sampling, geschichtete Stichproben und "Matrjoschka-sampling"

Wenn die theoriebezogene Repräsentativität einer Stichprobe durch ein kriteriengeleitetes Samplingverfahren sichergestellt werden soll, ist die Frage, *auf welche Weise die Kriterien für die Stichprobenziehung bestimmt werden können*, von zentraler Bedeutung.

Da qualitative Verfahren auf *Methodologien empirisch begründeter Theorienkonstruktion* beruhen, bei denen die Hypothesengenerierung und Theoriebildung auf der Basis empirischen Materials erfolgen soll, werden dort häufig *sequentielle Ziehungsverfahren* eingesetzt. Bei solchen Verfahren - in der Literatur als *theoretical sampling* (GLASER, STRAUSS 1967) oder *negative case analysis* beschrieben (LINDESMITH 1968; CRESSEY 1971) - werden auf der Basis einer sorgfältigen Analyse des zuerst gezogenen Falles Hypothesen entwickelt und danach auf dieser Basis weitere Fälle erhoben. Hierzu können einerseits Untersuchungseinheiten ausgewählt werden, die dieselben oder ähnliche Merkmalsausprägungen aufweisen wie die bislang untersuchten Fälle ("*minimizing differences*" (GLASER, STRAUSS 1967, S.55) oder "*Comparable-Case-Selection*" (LECOMPTE, PREISSLE 1993, S.78)) oder es können andererseits Fälle auf der Basis divergierender Merkmalsausprägungen ("*maximizing differences*", "*kontrastierende Fälle*") gesucht werden. Deren Analyse führt zu weiteren Hypothesen, die wiederum zur Ziehung neuer Fälle genutzt werden. Dieses Vorgehen, bei dem sich Datenauswertung und Stichproben-ziehung abwechseln, wird solange fortgeführt, bis sichergestellt scheint, daß durch die Untersuchung weiterer Fälle keine widersprechende

Gegenevidenz mehr gefunden wird (CRESSEY 1971) bzw. keine neuen Erkenntnisse mehr gewonnen werden können, d.h. eine "theoretische Sättigung" des Samples erreicht ist (GLASER, STRAUSS 1967).

Weil Fallanalyse und Stichprobenziehung bei diesem Vorgehen Hand in Hand gehen, verfügen die Untersucher zu Anfang über keinen Stichprobenplan, welcher ihnen Kriterien zur Ziehung des Samples oder die Stichprobengröße angibt. Dieses Verfahren erfordert also offene Untersuchungsdesigns, in denen weder Stichprobenpläne noch die Dauer des Erhebungsprozesses festgelegt sind. Es ist deswegen besonders dann geeignet, wenn die Untersucher nicht über Orientierungshypothesen über den untersuchten Gegenstandsbereich verfügen. Sein Anwendungsbereich liegt deshalb besonders auf dem Gebiet naturalistischer Feldforschung, ethnographischer Untersuchungen und explorativer Studien.

Kriteriengeleitete Stichprobenauswahlen können jedoch auch auf andere Weise realisiert werden. Dann wenn die Untersucher über Kenntnisse oder Arbeitshypothesen über relevante strukturelle Einflußfaktoren verfügen, kann durch eine a priori Definition von Auswahlmerkmalen sichergestellt werden, daß Träger bestimmter theoretisch relevanter Merkmalskombinationen im Sample vertreten sind und damit theoretisch bedeutsame Verzerrungen vermieden werden. Im Unterschied zum *theoretical sampling* sensu Glaser und Strauss werden bei diesem Vorgehen der kriteriengeleiteten Auswahl der Stichprobenumfang und die Ziehungskriterien vor der Erhebung festgelegt und die Daten erst nach der Erhebung analysiert. Eine solche Samplingstrategie wird in der qualitativen Forschungspraxis sehr häufig verwendet, in der Literatur jedoch nur selten beschrieben (so bei BRYMAN 1988, S. 136f.; FRETER, HOLLSTEIN, WERLE 1991).

Hinsichtlich der Auswahl der Untersuchungseinheiten kommen vier Erhebungstechniken in Betracht: Schichtungs-, Quotierungs- und Schneeballverfahren sowie die Auswahl aufgrund von Expertenurteilen. *Schneeballverfahren* und die *Auswahl auf der Basis von Expertenverfahren* (LECOMPTE, PREISSE 1993, S. 73, 76) können allerdings zu vollständig unkontrollierbaren Verzerrungen führen und sollten deshalb nur dann eingesetzt werden, wenn Angehörige der Zielpopulation anders nicht erreichbar sind. Da auch Quotierungsverfahren nur einen begrenzten Schutz gegen Verzerrungen bieten (vgl. SCHNELL, HILL, ESSER 1989, S. 279), ist die Ziehung einer *proportional* oder *dysproportional geschichteten Zufallsstichprobe* in jedem Fall als Methode der Wahl zu betrachten.

Bei der theoriegeleiteten Stichprobenziehung einer geschichteten Zufallsstichprobe müssen die Schichtungsmerkmale dabei *grundsätzlich* anhand *theoretischer Vorüberlegungen* bestimmt werden. Die im Folgenden dargestellten Samplingstrategien wurden vor dem Hintergrund der grundlegenden theoretischen Annahmen des Forschungsprogramms des Sfb 186 und der damit zusammenhängenden methodologischen Erfordernisse entwickelt. Die Untersuchung der Modernisierung von Lebensläufen bringt zwei unterschiedliche theoretische Perspektiven miteinander in Kontakt, die bislang verschiedenen methodologischen Paradigmen zugeordnet werden: einer *interpretativ-subjektbezogenen Perspektive*, die die individuellen Deutungsmuster und Handlungsorientierungen der Akteure in den Mittelpunkt stellt und eine *normativ-strukturelle* Perspektive, in deren Kontext soziales Handeln vorrangig als Produkt objektiver Sozialstrukturen verstanden wird. Einerseits müssen also jene *sozialstrukturellen Bedingungen* mit Hilfe quantitativer Methoden identifiziert werden, die die Handlungskontexte der Akteure ausmachen. Andererseits sollen mit Hilfe qualitativer Verfahren jene *Deutungsmuster und Handlungsorientierungen* systematisch in den Blick genommen werden, mit denen sie auf die strukturell vorgegebenen Situationslogiken und Handlungsoptionen reagieren. Bei der Erforschung von Deutungsmustern und Handlungsorientierungen der Akteure müssen dabei die jeweils unterschiedlichen sozialstrukturell bedingten Gelegenheitsstrukturen berücksichtigt werden. Durch eine kriteriengeleitete Stichprobenziehung soll dabei sichergestellt werden, daß die wesentlichen sozialstrukturellen Kontextbedingungen, die für das untersuchte Handlungsfeld relevant sind, bei der Auswahl von Untersuchungseinheiten für eine qualitative Untersuchung systematisch zur Geltung kommen. Vor der Stichprobenziehung müssen also die wesentlichen sozialstrukturellen Einflüsse, denen das Handeln der Akteure in dem untersuchten Feld unterliegt, identifiziert werden. Diese Einflußfaktoren gehen als Schichtungsmerkmale in den Stichprobenplan ein. Abhängig von dem Untersuchungsgegenstand und der jeweiligen theoretischen Perspektive bietet sich dabei eine von zwei Strategien an:

1. Die Schichtungsmerkmale werden aufgrund theoretischer Setzungen a priori festgelegt, etwa indem "Geschlecht", "Schichtzugehörigkeit" oder "Bildungsabschluß" (oder sonstige demographische Merkmale) als entscheidende Einflußgrößen bestimmt werden. Wird ein Einfluß dieser Variablen auf das untersuchte Handeln angenommen, so verlangt eine rationale Samplingstrategie, daß diese Merkmale in einem Stichprobenplan nach den Prinzipien der *bewußt heterogenen Auswahl* oder der *Auswahl nach*

Modalkategorien so miteinander kombiniert werden, daß Vertreter aller relevanten Merkmalskombinationen ausreichend im Sample vertreten sind.

2. Relevante Schichtungsmerkmale können jedoch auch auf der Basis quantitativer Voruntersuchungen entwickelt werden. Ein Beispiel hierfür bietet das Forschungsprojekt D3 zur Untersuchung von "Sozialhilfekarrieren". Auf der Basis quantitativer Untersuchungen gelang es, verschiedene Typen von Sozialhilfebezieher, genannt "transitorische Bezieher (Kurzzeitfälle)", "Mehrfachtransitorische", "Pendler", "Langzeitfälle" und "Ausbrecher" zu identifizieren. In der anschließenden qualitativen Untersuchung wurde geprüft, inwieweit sich diese Typen hinsichtlich ihrer Verarbeitungsformen des Sozialhilfebezugs unterscheiden. In diesem Fall kann eine Form der theoriegeleiteten Stichprobenziehung verwirklicht werden, die wir als *Matrjoschka-sampling* bezeichnen. Hierbei wird ein dysproportional geschichtetes qualitatives Sample aus einer größeren quantitativen Stichprobe so gezogen, daß - ähnlich wie bei der russischen Holzpuppe gleichen Namens - die kleine Stichprobe bestimmte Merkmale der größeren Stichprobe (manchmal allerdings in bewußt verzerrter Form) trägt.

Eine solche Form der Stichprobenziehung kann eine bestimmte Bandbreite sozialstruktureller Einflüsse abdecken, indem sichergestellt wird, daß alle theoretisch relevant erscheinenden Strukturmerkmale in der qualitativen Stichprobe in ausreichendem Umfang durch Einzelfälle vertreten sind. Mit Hilfe qualitativer Methoden kann dann untersucht werden, wie die Akteure ihre sozialstrukturell vorgegebenen Handlungsoptionen und -einschränkungen wahrnehmen und deuten, welche Handlungsziele sie unter diesen Bedingungen entwickeln und welche Mittel sie zur Erreichung dieser Ziele einsetzen.

Bei einem solchen Vorgehen wird das ursprünglich von Barton und Lazarsfeld empfohlene *Phasenmodell* für die Integration qualitativer und quantitativer Methoden (einer qualitativen pilot-study, die der Hypothesengenerierung dient, soll eine quantitative Hauptuntersuchung folgen) umgedreht. Vielmehr wird eine quantitative Studie, bei der gewissermaßen die Topographie der strukturellen Bedingungen des Handlungsfeldes kartographiert wird, zur *strategischen Plazierung* einer qualitativen Stichprobe verwendet.

Auch innerhalb einer so gezogenen qualitativen Stichprobe können - in Anlehnung an das Verfahren von Glaser und Strauss (1967) - Fälle für die Feinanalysen kontrastierend ausgewählt und ausgewertet werden. Durch die Berücksichtigung der theoretisch relevant erscheinenden Strukturmerkmale bei der Stichprobenziehung ist dabei bis zu einem gewissen Grad sichergestellt, daß keine unbekannten oder ungewollten Verzerrungen vorliegen und keine "weißen Flecken" bleiben: die Auswahl der Fälle findet hinsichtlich der quantitativ erhobenen Merkmale kontrolliert statt. Inwieweit etwa qualitativ erhobene Deutungsmuster oder Situationsdefinitionen auf die Populationen der jeweiligen Merkmalsträger übertragbar sind, hängt allerdings einerseits von deren Binnenvarianz und andererseits von der Anzahl der ausgewählten Fälle pro Gruppe ab; sind die Zellenbesetzungen innerhalb des qualitativen Stichprobenplans sehr klein, lassen sich die Homogenität und die Bandbreite von Deutungen innerhalb einer Gruppe von Merkmalsträgern natürlich oftmals nur unzureichend abschätzen. Bei sehr kleinen Fallzahlen ist daher nur schwerlich nachzuweisen, daß Deutungsmuster einzelner Akteure als repräsentativ für die von ihnen vertretene Gruppen gelten können. Das Ziel einer theoriegeleiteten Auswahl besteht allerdings nicht, wie schon erwähnt, darin, ein maßstabsgetreu verkleinertes Abbild einer Grundgesamtheit herzustellen, sondern darin, theoretisch bedeutsame Merkmalskombinationen bei der Stichprobenziehung zu berücksichtigen.

4.2.3 Kommunikative Validierung und qualitative Panels

Eine Methode, mit deren Hilfe jenen Bedrohungen der externen Validität begegnet werden kann, die von der spezifischen sozialen Struktur der Befragungssituation ausgehen (der *ad quem* Struktur, wie Gerhardt es nennt), stellt die wiederholte Befragung derselben Interviewpartner dar. Ein solches Vorgehen schlägt bspw. Zetterberg für die quantitative Befragungsmethodik insbesondere dann vor, wenn Faktenwissen abgefragt wird: *"When the interview method is employed to obtain factual information, cross checks can often be made (...) We can come back to the same respondent and ask him again (...)." (ZETTERBERG 1965, S. 123)*

Auch für die qualitative Forschung wurde von einigen Autoren ein solches Vorgehen unter dem Begriff der *kommunikativen Validierung* empfohlen. Hierbei handele es sich um ein *"methodisches Verfahren, sich der Gültigkeit einer Interpretation dadurch zu vergewissern, daß eine Einigung resp. Übereinstimmung über die Interpretation zwischen Interviewten und Interpreten hergestellt wird"* (KLÜVER 1979, 69). *"Unter 'kommunikativer Validierung' verstehen wir eine Form 'externer Validierung', bei der Interviewinterpretationen mit Betroffenen reflektiert werden."* (HEINZE, KLUSEMANN 1979, S. 195 f)

Das Verfahren, zu Beginn der achtziger Jahre in der Aktionsforschung entwickelt, kann sowohl bei der Datenerhebung als auch während der Auswertung des erhobenen Materials angewandt werden. Ziel ist ein Verständigungsprozeß zwischen Befragten und Forschern und Forscherinnen, der zu einer Validierung der Untersuchungsergebnisse führen soll. Hierbei wird die Zustimmung des 'Forschungssubjektes' als Gütekriterium eingeführt, da sie allein als die "Experten ihrer Lebenssituation" betrachtet werden.

Während mit dieser Methode ein forschungsethisch begrüßenswerter Anspruch erhoben wird - die "Forschungssubjekte" sollen stärker als bislang in der Sozialforschung üblich in den Prozeß der Theoriebildung einbezogen werden - so ist sie doch auch, was die externe Validität der Untersuchungsergebnisse angeht - mit nicht zu unterschätzenden Problemen behaftet:

- * Die Methode der kommunikativen Validierung basiert auf dem Postulat, daß Validität gleichbedeutend mit einer im Rahmen eines "herrschaftsfreien Diskurses" hergestellten Intersubjektivität sei. In den Interaktionen zwischen Forschern und Feldakteuren sowie zwischen einzelnen Feldakteuren ist

demgegenüber realistischerweise immer mit Asymetrien - etwa von Interessen oder Kommunikationskompetenzen - zu rechnen, die diese Annahme stark in Frage stellen. Kommunikative Validierung bedarf deshalb zumindest eines für alle Teilnehmer transparenten und geregelten Verfahrens, um Asymetrien zwar nicht aufzuheben, zumindest aber abzuschwächen und transparent zu gestalten.

- * Auch dort, wo ein solches geregeltes Verfahren existiert, bleibt die Frage, auf welcher *Ebene* Ergebnisse qualitativer Forschung von den Akteuren validiert werden können. Dies ist vor allem dort möglich, wo es um einzelne Fakten, Interpretationen einzelner Akteursdeutungen oder manifester Handlungsmotive geht. Sobald Interpretationen allerdings über die Ebene manifester Gehalte hinausgehen, wird die Zustimmung der Betroffenen als Validierungskriterium problematisch. Dies ist auch dann der Fall, wenn es sich bei den zu prüfenden Aussagen nicht mehr nur um die Sichtweisen einzelner Akteure handelt, sondern um Konstruktionen zweiter Ordnung, d.h. um Typen oder Modelle, die der Sozialforscher selber entwickelt hat. So äußert Terhart *"Zweifel am Dialog"*, wenn kommunikative Validierung als einziges oder auch nur zentrales Geltungskriterium angewandt wird: *"Sozialforschung wäre damit überflüssig geworden, weil ihre Erkenntnisse nie über das hinausgehen könnten und dürften, was gesellschaftlich Handelnde selbst schon von sich bzw. ihrem gesellschaftlichen Umfeld wissen bzw. zu wissen meinen."* (TERHART 1981, S.779).
- * Wenn jedoch nicht nur einzelne Verständnisfragen durch Rückfragen bei den Untersuchten geklärt werden sollen, sondern die auf der Basis des Datenmaterials erhobene Theorie, gelten die zur diskursiven Validierung angestellten Überlegungen analog: in stark heterogenen Gruppen sowie bei Forschungsthemen, die starke Kontroversen hervorrufen, ist ein Konsens über übergreifende Theorien kaum herstellbar. Es stellt sich damit z.B. die Frage, ob ein Dissens mit einem einzigen Mitglied der untersuchten Gruppe oder Population ausreicht, um eine Theorie über die Gruppe zu falsifizieren. In stark homogenen Gruppen besteht demgegenüber wiederum die Gefahr, daß ein Konsens - d.h. auch ein Konsens über falsche Deutungen - zu schnell und zu reibungslos herstellbar ist.

Das Konzept der kommunikativen Validierung läßt sich also kaum zum allgemeinen Modell der Validitätssicherung in der qualitativen Sozialforschung (bspw. auf der Basis

eines "konsensstheoretischen" Wahrheitsbegriffs (KVALE 1991)) ausweiten, um die Gültigkeit empirisch begründeter Theorien auszuweisen. Beschränkt man es jedoch auf die Strategie, Verständnisprobleme bei der Auswertung qualitativer Interviews durch Rückfragen bei den Interviewten zu klären (Lincoln und Guba sprechen hier weitaus pragmatischer von einem "*member check*") ist es dahingegen ein brauchbares Verfahren zur Sicherung von *Teilaspekten* externen Validität.

Durch die Konstruktion *qualitativer panels* können einige der erwähnten Beschränkungen der kommunikativen Validierung überwunden werden. Hierbei werden Mitglieder eines Samples mehrfach über einen längeren Zeitraum befragt, wobei weitergehende Strategien der Geltungsprüfung einsetzbar sind:

- * Das Vorgehen, Unklarheiten mit den Betroffenen abzuklären, könnte dabei durch ein Verfahren, in dem die Interaktionen zwischen Forschern und Feldakteuren stärker formalisiert und damit stärker kontrollierbar sind, ergänzt werden.
- * In Folgeerhebungen zeigt sich zudem die Stabilität oder Situationsgebundenheit bestimmter Themen, Sichtweisen und Handlungsstrategien. Zum Beispiel kann die Zuordnung von bestimmten Personen zu bestimmten Typen, Gruppen oder Clustern somit überprüft werden.
- * Hierauf aufbauend können auf der Grundlage erster Interpretationen Prognosen für Ereignisse oder Entwicklungen formuliert werden, die im weiteren Verlauf - während der Analyse von Folgeerhebungen - überprüft werden können. Auf der Grundlage dieses Verfahrens können dann qualitative Forschungsstrategien entwickelt werden, die eine hypothesengenerierende erste Phase und eine hypothesenprüfende zweite Phase umfassen.

Auf diese Weise ließe sich die "kommunikative Validierung" (bzw. der "member check") mit der Aufstellung falsifizierbarer Verlaufshypothesen verbinden. Die grundlegende *ad quem* Struktur von Interviewsituationen wird hiermit jedoch nicht aufgehoben, in den meisten Fällen wahrscheinlich nicht einmal aufgebrochen. Es ist zu befürchten, daß Antworttendenzen wie *soziale Erwünschtheit* auch in Wiederholungsbefragungen wirksam werden, so daß hinsichtlich dieser Bedrohung der externen Validität nach weiteren Strategien der Geltungssicherung gesucht werden muß.

4.2.4 Das Konzept der "Triangulation"

Zur Vermeidung bzw. zur Korrektur der mit der *ad quem* Struktur von Interviewsituationen zusammenhängenden Fehler und Fehleinschätzungen kann auf eine Strategie der Validitätssicherung zurückgegriffen werden, die Denzin unter dem Begriff der "Triangulation" gefaßt hat¹⁵. Hierbei ist für die oben beschriebene Fragestellung insbesondere die sog. *across-methods-triangulation* (DENZIN 1977, S. 308 ff) von Bedeutung:

Hierbei werden zur Untersuchung *eines Gegenstandes* Daten und Modelle herangezogen, die *auf unterschiedlichen Methoden* der Erhebung und Analyse basieren. Die *Konvergenz* der Aussagen oder Modelle könne sodann als Validierungskriterium genutzt werden, da zwar jeder einzelnen Methode spezifische Biases und *threats for validity* innewohnen, diese durch Verwendung unterschiedlicher Methoden jedoch ausgeglichen würden. Denzin nimmt hier auf Webb Bezug, der fordert: "*Ideally, we should like to converge data from several different data classes, as well as converge with multiple variants from within a single class.*" (WEBB ET AL. 1966, S. 35).

Ein Beispiel für eine solche Vorgehensweise findet sich in den Arbeiten des Projekts D3 des Sfb 186 über Sozialhilfekarrieren. In diesem Rahmen wurden sowohl qualitative, leitfadengestützte Interviews mit Sozialhilfeempfängern durchgeführt als auch quantitative Daten auf der Grundlage von Sozialhilfeakten erhoben. Da die qualitative Stichprobe eine Teilmenge der quantitativen darstellte, konnten Aussagen in den Interviews mit Aufzeichnungen in den entsprechenden Akten abgeglichen werden. Hierbei zeigte sich, daß bei einer Gruppe von Interviewten, die eine stark

¹⁵ Ausführlicher wird dieser Ansatz diskutiert in Prein, Kelle, Kluge 1993.

diskontinuierliche Sozialhilfekarriere aufwies, die im Interview dargestellte Abfolge der unterschiedlichen Bezugsphasen in vielen Fällen nicht mit den Daten aus den Sozialhilfeakten deckungsgleich war. In diesem Fall wurde offensichtlich, daß die leitfadengestützten Interviews *in bezug auf die faktischen zeitlichen Rahmenbedingungen des Sozialhilfebezugs* nicht in jedem Fall valide und zuverlässige Ergebnisse lieferten, da die Schwerpunkte dieser Erhebung auf den subjektiven Relevanzstrukturen und Verarbeitungsmustern lagen. Ohne den Vergleich mit den quantitativen Daten hätten diese Verzerrungen z.T. zu schwerwiegenden Fehlinterpretationen führen können. Hier zeigt sich deutlich die Berechtigung von Gerhardts Kritik (vgl. GERHARDT 1985, S. 231), wonach die Narrativität von Interviewmaterial kein Garant für deren Validität hinsichtlich aller möglichen Aspekte darstellt.

Der Denzin'sche Ansatz der Methodentriangulation stellt somit in bestimmten Fällen eine Vorgehensweise dar, mit deren Hilfe unkontrollierte Einflüsse der Untersuchungssituation bzw. der Erhebungsmethode aufgedeckt und korrigiert werden können. Dies setzt allerdings voraus, daß alle eingesetzten Methoden sich auf den gleichen Erkenntnisgegenstand beziehen. Nur unter dieser Voraussetzung ist die Konvergenz von Ergebnissen ein sinnvolles Validierungskriterium. Daß dies nicht immer der Fall ist und daß somit die Methodentriangulation weniger unproblematisch ist, als Denzin dies darstellt, ist in der Vergangenheit vielfach betont worden (vgl. etwa FIELDING, FIELDING 1986; LAMNEK 1988; FLICK 1991; PREIN, KELLE, KLUGE 1993). Bei diesen Kritiken steht die Feststellung im Vordergrund, daß Untersuchungsmethoden stets auch der Ausdruck bestimmter theoretischer Grundannahmen über das Erkenntnisobjekt beinhalten und somit an dessen Konstitution mitbeteiligt sind. Konkret kann dies dazu führen, daß bei der Untersuchung desselben Gegenstandsbereichs mit unterschiedlichen Methoden jeweils unterschiedliche Aspekte in den Vordergrund gerückt werden - und damit die Untersuchungsergebnisse sich nicht *decken*, sondern eher wie Teile eines Puzzles *ergänzen*. Als Beispiel für eine solche Komplementarität von Forschungsergebnissen aus der Forschungspraxis läßt sich eine Untersuchung des Sfb-Teilprojekts B1 über Erwerbsverläufe von Frauen anführen (ERZBERGER 1993). Eine quantitative Erhebung zeigte, daß eine Korrelation zwischen den Arbeitsmarktchancen des Erstberufs und den Erwerbsverläufen der Frauen bestand: Frauen, deren erlernter Erstberuf gute Arbeitsmarktchancen hatte, wiesen in der Regel längere Erwerbsphasen in ihrem Lebenslauf auf als Frauen mit einem Erstberuf mit hohen Arbeitsmarktrisiken. In qualitativen Interviews hingegen nannten Frauen aus der gleichen Stichprobe fast ausschließlich familienbedingte Gründe für einen Wechsel der

Erwerbsstatus. Qualitativ und quantitativ erhobene Daten führten also nicht zu konvergierenden Ergebnissen. Die vermeintlich divergenten Ergebnisse konnten also nicht als falsifizierende Instanz gewertet werden, sondern erwiesen sich im Rahmen eines *übergreifenden theoretischen Ansatzes* als *komplementär*, mit dem gleichermaßen die statistisch nachgewiesene Bedeutung des Erstberufs als auch die subjektiven Relevanzsetzungen der Befragten erklärt werden konnten: Berufe mit hohen Arbeitsmarktchancen waren als Ressourcen zu werten, die die Frauen in innerfamiliäre Auseinandersetzungen über ihre Erwerbsbeteiligung stillschweigend mit einbrachten, während die Geltung des normativen Prinzips der männlichen Ernährerrolle in der Familie subjektiv für die befragten Frauen weiterhin Geltung behielt.

Aus den oben angeführten Beispielen wird einerseits deutlich, daß zur Abwehr von Validitätsbedrohungen, die aufgrund von Einflüssen der Erhebungssituation oder -methode zu Methodenartefakten führen *in bestimmten Situationen* auf den Ansatz der Triangulation Bezug genommen werden kann. Dies ist vor allem dann der Fall, wenn sich die verschiedenen Methoden *auf den gleichen Gegenstand* beziehen (oder die Gegenstandsbereiche der verschiedenen Methoden sich zumindest in einem hohen Maß überschneiden) und damit überhaupt nur zu konvergenten Ergebnissen führen können. Ebenso wenig wie das Konzept der kommunikativen Validierung läßt sich dieser Ansatz allerdings zu einer allgemeinen Validierungsstrategie ausweiten, der auf jede Forschungsfrage in jeder Situation anwendbar ist. In vielen Fällen beziehen sich unterschiedliche Methoden auf einen unterschiedlichen Gegenstandsbereich und können deshalb keine *konvergenten Ergebnisse* liefern, die sich zur gegenseitigen Validierung nutzen lassen, sondern *komplementäre Ergebnisse*, die zusammengenommen ein stimmiges Gesamtbild ergeben.

5 Literatur

- AGAR, MICHAEL (1986): *Speaking of Ethnography*. Beverly Hills: Sage Publications.
- ALBERT, HANS (1980): Theorie und Prognose in den Sozialwissenschaften. In: TOPITSCH, ERNST (Hg.): *Logik der Sozialwissenschaften*. S. 126-143.
- APEL, KARL-OTTO (1979): *Die Erklären:Verstehen-Kontroverse in transzendental-pragmatischer Sicht*. Frankfurt/M.: Suhrkamp.
- APEL, KARL-OTTO; MANNINEN, JUHA; TUOMELA, RAIMO (Hg.) (1978): *Neue Versuche über Erklären und Verstehen*. Frankfurt/M.: Suhrkamp.
- BAIN, R. (1929): The Validity of Life Histories and Diaries. In: *Journal of Educational Sociology*, 3, S. 150-164.
- BARTON, ALLEN H.; LAZARSELD PAUL F. (1955): Some Functions of Qualitative Analysis in Social Research. In: *Frankfurter Beiträge zur Soziologie*. Frankfurt: Europäische Verlagsanstalt. S. 321-361.
- BARTON, ALLEN H.; LAZARSELD, PAUL F. (1984): *Einige Funktionen von qualitativer Analyse in der Sozialforschung*. In: HOPF, CHRISTEL; WEINGARTEN, ELMAR (Hg.): *Qualitative Sozialforschung*. Stuttgart: Klett-Cotta, S. 41-89 (1. Aufl.: 1979). (Quelle für die dt. Übersetzung: Barton, Lazarsfeld 1955).
- BENARD, CHERYL; SCHLAFFER, EDITH (1991): *Sagt uns, wo die Väter sind. Von Arbeitssucht und Fahnenflucht des zweiten Elternteils*. Reinbek: Rowohlt.
- BERGMANN, J. (1985): Flüchtigkeit und methodische Fixierung sozialer Wirklichkeit. In: BONß, W.; HARTMANN, H. (Hg.): *Entzauberte Wissenschaft*. Göttingen. S. 299-320.
- BERTAUX, DANIEL (1981): Introduction. In: Ders.: *Biography and Society. The Life History Approach in the Social Sciences*. London and Beverly Hills, S. 5 - 15.
- BOHNSACK, RALF (1991): *Rekonstruktive Sozialforschung. Einführung in Methodologie und Praxis qualitativer Forschung*. Opladen: Leske und Budrich.
- BRYMAN, ALAN (1988): *Quantity and Quality in Social Research*. London, New York: Routledge.
- CICOUREL, AARON V. (1974): *Methode und Messung in der Soziologie*. Frankfurt a. M.: Suhrkamp.
- COOK, THOMAS D.; CAMPBELL, DONALD T. (1979): *Quasi-Experimentation. Design & Analysis Issues for Field Settings*. Boston: Houghton Mifflin Company.
- CORBIN, JULIET; STRAUSS, ANSELM (1990): Grounded Theory Research: Procedures, Canons and Evaluative Criteria. In: *Zeitschrift für Soziologie*, 19, S. 418-427.

- CRESSEY, DONALD R. (1950): The Criminal Violation of Financial Trust. In: *American Sociological Review*, 15, S. 738-743.
- CRESSEY, DONALD R. (1971): *Other People's Money. A Study in the Social Psychology of Embezzlement*. Belmont: Wadsworth (erstmal erschienen 1953).
- CRONBACH, L. J.; MEEHL, P. E. (1955): Construct Validity in Psychological Tests. In: *Psychological Bulletin*, 52, S. 281-302.
- DANNEBERG, LUTZ (1989): *Methodologien. Struktur, Aufbau und Evaluation*. Berlin: Duncker und Humblot.
- DENZIN, NORMAN K. (1977): *The Research Act. A Theoretical Introduction to Sociological Methods*. New York etc.: McGraw Hill Book Company.
- DILTHEY, W. (1924): *Gesammelte Schriften. Die geistige Welt. Einleitung in die Philosophie des Lebens*, 5. Bd. Leipzig, Berlin (darin: Die Entstehung der Hermeneutik (1900)).
- EISNER, ELLIOT W. (1981): On the Differences between Scientific and Artistic Approaches to Qualitative Research. In: *Educational Researcher*, 10, S. 5-9.
- ERZBERGER, CHRISTIAN (1993): *Erwerbsarbeit im Eheleben. Männlicher und weiblicher Erwerbsverlauf zwischen Dependenz und Unabhängigkeit*. Bremen: Arbeitspapiere des Sfb 186, Nr. 16.
- ESSER, HARTMUT (1987): Zum Verhältnis von qualitativen und quantitativen Methoden in der Sozialforschung, oder: Über den Nutzen methodologischer Regeln bei der Diskussion von Scheinkontroversen. In: Voges, Wolfgang (Hrsg.) (1987): *Methoden der Biographie- und Lebenslaufforschung*. Opladen: Westdeutscher Verlag. S. 87-101.
- ESSER, HARTMUT (1989): Verfällt die "soziologische Methode"? In: *Soziale Welt*, 40, S. 57-75.
- ESSER, HARTMUT (1990): "Habits", "Frames" und "Rational Choice". Die Reichweite der Theorien der rationalen Wahl (am Beispiel der Erklärung des Befragtenverhaltens). In: *Zeitschrift für Soziologie*, 19, S. 231-247.
- ESSER, HARTMUT (1991): *Alltagshandeln und Verstehen. Zum Verhältnis von erklärender und verstehender Soziologie am Beispiel von Alfred Schütz und "Rational Choice"*. Tübingen: J.C.B. Mohr.
- FIELDING, NIGEL G.; FIELDING, JANE L. (1986): *Linking Data. Qualitative Research Methods*, Volume 4. Beverly Hills etc.: Sage.
- FILSTEAD, WILLIAM J. (1979): Soziale Welten aus erster Hand. In: GERDES, KLAUS (Hg.): *Explorative Sozialforschung*. Stuttgart: Enke. S. 29-40.
- FISCHER, KLAUS (1983): Rationale Heuristik. In: *Zeitschrift für allgemeine Wissenschaftstheorie*, 14, S. 234-272.

- FLECK, CHRISTIAN (1992): Vom "Neuanfang" zur Disziplin? Überlegungen zur deutschsprachigen qualitativen Sozialforschung anlässlich einiger neuer Lehrbücher. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, Jg. 44, Heft 4, S. 747-765.
- FLICK, UWE (1987): Methodenangemessene Gütekriterien in der qualitativ-interpretativen Forschung. In: BERGOLD, JARG B.; FLICK, UWE (Hg.): *Ein-Sichten: Zugänge zur Sicht des Subjekts mittels qualitativer Forschung*. Tübingen: DGVT (Forum für Verhaltenstherapie und soziale Praxis Nr.14). S. 247-262.
- FLICK, Uwe (1991): Triangulation. In: FLICK, UWE ET AL. (Hg.): *Handbuch Qualitative Sozialforschung. Grundlagen, Konzepte, Methoden und Anwendungen*. München und Weinheim: Psychologie Verlags Union. S. 432-424.
- FLICK, UWE (1992): Triangulation Revisited: Strategy of Validation or Alternative? In: *Journal for The Theory of Social Behaviour*. Vol. 22, Nr. 2, Juni 1992, S. 175-197.
- FLICK, UWE (1993): Triangulation Revisited: Strategy of Validation or Alternative. In: *Journal for the Theory of Social Behaviour*, 22, S. 175-197.
- FLICK, UWE; KARDORFF, ERNST V.; KEUPP, HEINER; ROSENSTIEL, LUTZ V.; WOLFF, STEPHAN (Hg.) (1991): *Handbuch qualitative Sozialforschung: Grundlagen, Konzepte, Methoden und Anwendungen*. München: Psychologie Verlags Union.
- FREEMAN, DEREK (1983): *Margaret Mead and Samoa: The Making and Unmaking of an Anthropological Myth*. Cambridge, MA: Harvard University Press.
- FRETER, HANS-JÜRGEN; HOLLSTEIN, BETINA; WERLE, Markus (1991): Integration qualitativer und quantitativer Verfahrensweisen - Methodologie und Forschungspraxis. In: *ZUMA-Nachrichten*, Nr. 29, November 1991, S. 98-114.
- FRIEDRICHS, JÜRGEN (1980): *Methoden empirischer Sozialforschung*. Opladen: Westdeutscher Verlag (erstmal erschienen 1973).
- FRIEDRICHS, JÜRGEN; LÜDTKE, HANS (1973): *Teilnehmende Beobachtung. Einführung in die sozialwissenschaftliche Feldforschung*. Weinheim und Basel: Beltz Verlag.
- GERDES, KLAUS (1979) (Hg.): *Explorative Sozialforschung*. Stuttgart: Enke.
- GERHARDT, UTA (1985): Erzählenden und Hypothesenkonstruktion: Überlegungen zum Gültigkeitsproblem in der biographischen Sozialforschung. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 37, S. 230-256.
- GERHARDT, UTA (1986): Verstehende Strukturanalyse: Die Konstruktion von Idealtypen als Analyseschritt bei der Auswertung qualitativer Forschungsmaterialien. In: SOEFFNER, HANS-GEORG (Hg.): *Sozialstruktur und soziale Typik*. Frankfurt/M., New York: Campus. S. 31-83.
- GERHARDT, UTA (1991): Typenbildung. In: FLICK, UWE; KARDORFF, ERNST V.; KEUPP, HEINER; ROSENSTIEL, LUTZ V.; WOLFF, STEPHAN (Hg.): *Handbuch qualitative*

Sozialforschung: Grundlagen, Konzepte, Methoden und Anwendungen. München: Psychologie Verlags Union. S. 435-439.

GIDDENS, ANTHONY (1984): *Interpretative Soziologie. Eine kritische Einführung.* Frankfurt, New York: Campus.

GLASER, BARNEY G. (1965): The Constant Comparative Method of Qualitative Analysis. In: *Social Problems*, 12, S. 436-445.

GLASER, BARNEY G. (1978): *Theoretical Sensitivity. Advances in the Methodology of Grounded Theory.* Mill Valley: The Sociology Press.

GLASER, BARNEY G.; STRAUSS, ANSELM L. (1967): *The Discovery of Grounded Theory: Strategies for Qualitative Research.* New York: Aldine de Gruyter.

GLASER, BARNEY G.; STRAUSS, ANSELM L. (1974): *Interaktion mit Sterbenden.* Göttingen: Vandenhoeck & Ruprecht (erstmal 1968 erschienen unter dem Titel "Time for Dying" (Chicago: Aldine)).

GUTTMAN, LOUIS (1941): Mathematical and Tabulation Techniques. In: SOCIAL SCIENCE RESEARCH COUNCIL (Hg.): *The Prediction of Personal Adjustment.* New York.

GUTTMAN, LOUIS (1950): The Problem of Attitude and Opinion Measurement. In: STOUFFER, SAMUEL (Hg.): *Measurement and Prediction.* Princeton: Princeton University Press, S. 46-59.

HAMMERSLEY, MARTYN (1992): Some Reflections about Ethnography and Validity. In: *Qualitative Studies in Education*, 5, S. 195-203.

HANSON, NORWOOD RUSSELL (1965): *Patterns of Discovery. An Inquiry Into the Conceptual Foundations of Science.* Cambridge: Cambridge University Press. (erstmal erschienen 1958).

HEINZE, THOMAS; KLUSEMANN, HANS W. (1979): Ein biographisches Interview als Zugang zu einer Bildungsgeschichte. In: BAACKE, D.; SCHULZE, TH. (Hg.): *Aus Geschichten lernen.* München: Juventa, S.182-225.

- HEINZE, THOMAS; KLUSEMANN, HANS W. (1980): Versuch einer sozialwissenschaftlichen Paraphrasierung am Beispiel des Ausschnittes einer Bildungsgeschichte. In: HEINZE, THOMAS; KLUSEMANN, HANS W.; SOEFFNER, HANS GEORG (Hg.): *Interpretationen einer Bildungsgeschichte. Überlegungen zur sozialwissenschaftlichen Hermeneutik*. Bensheim: Päd. extra. S. 97-152.
- HEINZE, THOMAS; THIEMANN, FRIEDRICH (1982): Kommunikative Validierung und das Problem der Geltungsbegründung. Bemerkungen zum Beitrag von E. Terhart. In: *Zeitschrift für Pädagogik*, 28, S. 635-642.
- KELLE, UDO (1990): *Computerunterstützte Auswertung qualitativer Daten*. Bremen: Arbeitspapiere des Sfb 186, Nr. 11.
- KELLE, UDO (1992): *Empirisch begründete Theoriebildung*. Ein Beitrag zur Logik und Methodologie interpretativer Sozialforschung. Dissertationsschrift. Universität Bremen.
- KIRK, JEROME; MILLER, MARC L. (1986): *Reliability and Validity in Qualitative Research* (Qualitative Research Methods, Vol. 1). London: Sage Publications.
- KLEINING, G. (1982): Umriß zu einer Methodologie qualitativer Sozialforschung. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 34, S. 224-253.
- KLÜVER, J. (1979): Kommunikative Validierung - einige vorbereitende Bemerkungen zum Projekt "Lebensweltanalyse von Fernstudenten". In: HEINZE, THOMAS (Hg.): *Lebensweltanalyse von Fernstudenten. Theoretische und methodologische Überlegungen zum Typus hermeneutisch-lebensgeschichtlicher Forschung*. Fernuniversität Hagen, S. 68-84.
- KOCK, BIRGIT; WITZEL, ANDREAS (1993): *Berufsbiographische Gestaltungsprinzipien. Theoretische und methodische Grundlagen*. Arbeitspapier des Sfb 186, Nr. 22, Universität Bremen.
- KRIZ, JÜRGEN; LISCH, RALF (1988): *Methodenlexikon für Mediziner, Psychologen, Soziologen*. München, Weinheim: Psychologie Verlags Union.
- KROMREY, HELMUT (1990): *Empirische Sozialforschung: Modelle und Methoden der Datenerhebung und Datenauswertung*. Opladen: Leske und Budrich.
- KÜCHLER, MANFRED (1983): "Qualitative" Sozialforschung - ein neuer Königsweg?. In: GARZ, DETLEF; KRAIMER, KLAUS (Hg.): *Brauchen wir andere Forschungsmethoden. Beiträge zur Diskussion interpretativer Verfahren*. Frankfurt/Main: Athenäum. S. 9-31.
- KVALE, STEINAR (HG.) (1989): *Issues of Validity in Qualitative Research*. Lund: Studentlitteratur.

- KVALE, STEINAR (1991): Validierung: Von der Beobachtung zu Kommunikation und Handeln. In: FLICK, UWE et al. (Hg.): *Handbuch Qualitative Sozialforschung*. München und Weinheim: Psychologie Verlags Union. S. 427-431.
- KVALE, STEINAR (1991): Validierung: Von der Beobachtung zu Kommunikation und Handeln. In: FLICK, UWE et al. (Hg.): *Handbuch Qualitative Sozialforschung*. München und Weinheim: Psychologie Verlags Union. S.427-431.
- LAKATOS, IMRE (1982): *Die Methodologie der wissenschaftlichen Forschungsprogramme*. Philosophische Schriften, Bd.1 Wiesbaden: Vieweg.
- LAMNEK, SIEGFRIED (1988): *Qualitative Sozialforschung. Bd. 1: Methodologie*. München und Weinheim: Psychologie Verlags Union.
- LAUDAN, LARRY (1977): *Progress and its Problems. Towards a Theory of Scientific Growth*. London and Henley: Routledge & Kegan Paul.
- LECHLER, PETER (1982): Kommunikative Validierung. In: HUBER, GÜNTER L.; MANDL, HEINZ (Hg.): *Verbale Daten. Eine Einführung in Erhebung und Auswertung*. Weinheim und Basel: Beltz.
- LECOMPTE, MARGARET D.; PREISSE, JUDITH (1993): *Ethnography and Qualitative Design in Educational Research*. San Diego: Academic Press.
- LEE, RAYMOND M.; FIELDING, NIGEL G. (ed.) (1991): *Using Computers in Qualitative Research*. London pp.: Sage.
- LINCOLN, YVONNA S.; GUBA, EGON G. (1985): *Naturalistic Inquiry*. Beverly Hills pp.: Sage Publications.
- LINDESMITH, ALFRED R. (1968): *Addiction and Opiates*. Chicago: Aldine (erstmal erschienen 1947).
- LUNDBERG, G. A. (1942): *Social Research: A Study in Methods of Gathering Data*. New York: Longmans, Green (erstmal erschienen 1929).
- MATT, EDUARD (1992): *Über ethnographische Beschreibungen oder Die Kunst der Konstruktion sozialer Wirklichkeit*. Gießen: Dissertation. Unveröffentl. Manuskript.
- MAYNTZ, RENATE; HOLM, KURT; HÜBNER, PETER (1969): *Einführung in die Methoden der empirischen Sozialforschung*. Köln: Westdeutscher Verlag.
- MAYRING, PHILLIP (1985): Qualitative Inhaltsanalyse. In: JÜTTEMANN, GERD (Hg.): *Qualitative Forschung in der Psychologie. Grundfragen, Verfahrensweisen, Anwendungsfelder*. Weinheim, Basel: Beltz.
- MAYRING, PHILLIP (1988): *Die qualitative Wende. Arbeiten zur qualitativen Forschung*. Augsburg: Augsburger Berichte zur Entwicklungspsychologie und Pädagogischen Psychologie 32.
- MAYRING, PHILLIP (1990): *Einführung in die qualitative Sozialforschung*. München: Psychologie Verlags Union.

- MEAD, MARGARET (1978): *Jugend und Sexualität in primitiven Gesellschaften. Bd. 1: Kindheit und Jugend in Samoa*. München: Deutscher Taschenbuch Verlag (erstmalig erschienen 1928 unter dem Titel "*Coming of Age in Samoa*" im Verlag William Morrow, New York).
- MILES, MATTHEW B.; HUBERMAN, A. MICHAEL (1993): *Qualitative Data Analysis: A Sourcebook of New Methods*. Beverly Hills, CA: Sage (erstmalig erschienen 1984).
- NICKLES, THOMAS (Hg.) (1980): *Scientific Discovery, Logic and Rationality* (Boston Studies in the Philosophy of Science, Vol. LVI). Reidel: Dordrecht
- NICKLES, THOMAS (1985): Beyond Divorce: Current Status of the Discovery Debate. In: *Philosophy of Science*, 52, S. 177-206.
- NICKLES, THOMAS (1990): Discovery Logics. In: *Philosophica*, 45, S. 732.
- OEVERMANN, ULRICH; ALLERT, TILMANN; KONAU, ELISABETH; KRAMBECK, JÜRGEN (1979): Die Methodologie einer 'objektiven Hermeneutik' und ihre allgemeine forschungslogische Bedeutung in den Sozialwissenschaften. In: SOEFFNER, HANS-GEORG: (Hg.): *Interpretative Verfahren in den Sozial- und Textwissenschaften*. Stuttgart: Metzler. S. 352-434.
- OPP, KARL-DIETER (1976): *Methodologie der Sozialwissenschaften. Einführung in Probleme ihrer Theoriebildung*. Reinbek: Rowohlt.
- OPP, KARL-DIETER (1987) : Wissenschaftstheoretische Grundlagen der empirischen Sozialforschung. In: ROTH, ERWIN (Hg.): *Sozialwissenschaftliche Methoden: Lehr- und Handbuch für Forschung und Praxis*. München, Wien: Oldenbourg. S. 47-71.
- POPPER, KARL R. (1989): *Logik der Forschung*. 9. verbesserte Auflage. Tübingen: J.C.B. Mohr (Paul Siebeck) (erstmalig erschienen 1934).
- PREIN, GERALD; KELLE, UDO; KLUGE, SUSANN (1993): *Strategien zur Integration quantitativer und qualitativer Auswertungsverfahren*. Bremen: Arbeitspapiere des Sfb 186, Nr. 19.
- REICHERTZ, JO (1986): *Probleme qualitativer Sozialforschung*. Frankfurt, New York: Campus.
- REICHERTZ, JO (1991): *Aufklärungsarbeit. Kriminalpolizisten und teilnehmende Beobachter bei der Arbeit*. Stuttgart: Enke.
- ROBINSON, W. S. (1951): The Logical Structure of Analytic Induction. In: *American Sociological Review*, 16, S. 812-818.
- ROBINSON, W. S. (1952): Rejoinder to Comments on "The Logical Structure of Analytic Induction". In: *American Sociological Review*, 17, S. 494.
- SAHNER, HEINZ (1989): *Schließende Statistik*. Stuttgart: Teubner.

- SCHNELL, RAINER; HILL, PAUL B.; ESSER, ELKE (1989): *Methoden der empirischen Sozialforschung*. München, Wien: Oldenbourg.
- SMITH, J. K. (1984): The Problem of Criteria for Judging Interpretive Inquiry. In: *Educational Evaluation and Policy Analysis*, 6, S. 379-391.
- SMITH, J. K. (1990): Alternative Research Paradigms and the Problem of Criteria. In: GUBA, EGON (Hg.): *The Paradigm Dialog*. Newbury Park: Sage Publications.
- SMITH, MARY LEE; GLASS, GENE V. (1987): *Research and Evaluation in Education and the Social Sciences*. Englewood Cliffs: Prentice Hall.
- SPÖHRING, WALTER (1989): *Qualitative Sozialforschung*. Teubner: Stuttgart.
- STEGMÜLLER, WOLFGANG (1974): *Wissenschaftliche Erklärung und Begründung. Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie*, Bd. 1. Berlin, Heidelberg, New York: Springer.
- TERHART, EWALD (1981): Intuition - Interpretation - Argumentation. Zum Problem der Geltungsbegründung von Interpretationen. In: *Zeitschrift für Pädagogik*, 27, S. 769-793.
- TESCH, RENATA (1990): *Qualitative Research. Analysis Types & Software Tools*. New York, Philadelphia, London: The Falmer Press.
- TESCH, RENATE (1992): Verfahren der computerunterstützten Analyse. In: HUBER, GÜNTER L. (HG): *Qualitative Analyse. Computereinsatz in der Sozialforschung*. München, Wien: Oldenbourg, S. 43-69.
- TREVOR-ROPPER, HUGH (1973): Re-inventing Hitler. *Sunday Times*, 18. Februar 1973.
- VOLMERG, UTE (1983): Validität im interpretativen Paradigma. Dargestellt an der Konstruktion qualitativer Erhebungsverfahren. In: ZEDLER, PETER; MOSER, HEINZ (Hg.): *Aspekte qualitativer Sozialforschung*. Opladen: Leske und Budrich. S. 124-143.
- WEBB, EUGENE J.; CAMPBELL, DONALD T.; SCHWARTZ, RICHARD D.; SECHREST, LEE (1966): *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago: Rand McNally.
- WEINGARTEN, ELMAR; SACK, FRITZ (1976): Ethnomethodologie. Die methodische Konstruktion der Realität. In: WEINGARTEN, ELMAR; SACK, FRITZ; SCHENKEIN, JIM (Hg.): *Ethnomethodologie*. Beiträge zu einer Soziologie des Alltagshandelns. Frankfurt a. M.: Suhrkamp. S. 7-26.
- WEINGARTEN, ELMAR; SACK, FRITZ; SCHENKEIN, JIM (1976) (HG.): *Ethnomethodologie*. Beiträge zu einer Soziologie des Alltagshandelns. Frankfurt a. M.: Suhrkamp.
- WILSON, THOMAS P. (1981): Theorien der Interaktion und Modelle soziologischer Erklärung. In: ARBEITSGRUPPE BIELEFELDER SOZIOLOGEN (Hg.): *Alltagswissen, Interaktion und gesellschaftliche Wirklichkeit*. Opladen: Westdeutscher Verlag (5. Auflage). S. 54-79.

ZETTERBERG, HANS L. (1965): *On Theory and Verification in Sociology*.
Totowa: The Bedminster Press (erstmalig erschienen 1953).